

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 06:59:35 ; Search time 59 Seconds
(without alignments)
1743.174 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 NM5GQKRLVMWAGTGGHV.....RVANEVSRVARELHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1841	97.9	364	5	Aau99816 Escherich
2	1803	95.9	355	4	Aau34438 E. coli c
3	1803	95.9	355	4	Ag98406 Escherich
4	1803	95.9	355	6	Au28491 Protein e
5	1687	89.7	355	6	Au47265 Protein e
6	1677	89.2	355	4	Au38485 Salmonell
7	1677	89.2	355	6	Au48097 Protein e
8	1604	85.3	348	6	Au31657 Protein e
9	1549.5	82.4	348	6	Au45116 Protein e
10	1471	78.2	356	6	Au50002 Protein e
11	1431	76.1	356	6	Am68867 Phototrab
12	1345.5	71.5	360	6	Au40887 Protein e
13	1323	70.3	281	6	Au27914 Protein e
14	1154	61.4	354	6	Au49532 Protein e
15	1123	59.7	354	6	Au38975 Protein e
16	1064.5	56.6	351	4	Au35596 Haemophil
17	1064.5	56.6	351	6	Au30498 Protein e
18	1060	56.4	351	5	Au99817 Haemophil
19	871	46.3	367	6	Au22301 Protein e
20	849	45.1	372	6	Au21174 Protein e
21	791	42.1	357	6	Au23450 Protein e
22	772.5	41.1	355	6	Au38137 Protein e
23	764.5	40.6	355	3	Aay74403 Neisseria
24	759.5	40.4	355	3	Aay74401 Neisseria
25	759.5	40.4	355	6	Abp80549 N. gonorr

26	759.5	40.4	355	6	ABU37201 Protein e
27	719.5	38.3	357	4	AAU36416 Pseudomon
28	719.5	38.3	357	6	ABU38727 Protein e
29	717	38.1	356	6	ABU41785 Protein e
30	696.5	37.0	367	6	ABU35148 Protein e
31	684	36.4	375	6	ABU40398 Protein e
32	680.5	36.2	366	6	ADA33749 Acinetoba
33	672.5	35.8	385	6	ABU16742 Protein e
34	630.5	33.5	335	3	AAY74402 Neisseria
35	535	28.4	408	6	ABU34101 Protein e
36	534	28.4	250	6	ABU19873 Protein e
37	510.5	27.1	407	6	ABU35864 Protein e
38	506	26.9	410	6	AAAG1172 Mycobacte
39	506	26.9	410	6	ABU15855 Mycobacte
40	506	26.9	410	6	ABU36727 Protein e
41	506	26.9	410	6	ABU34407 Protein e
42	488.5	26.0	359	6	ABU26097 Protein e
43	487	25.9	372	4	AAAG2109 C glutami
44	479	25.5	380	6	ABU20900 Protein e
45	477.5	25.4	409	5	AAU99823 Mycobacte
46	470.5	25.0	409	6	ABU25498 Protein e
47	460	24.5	383	6	ABU17984 Protein e
48	451.5	24.0	386	4	AAU60464 Propionib
49	451.5	24.0	386	6	ABM56983 Propionib
50	435	23.1	357	5	ABP54935 Lactobac
51	431.5	22.9	393	5	ABP66056 Bifidobac
52	423.5	22.5	366	6	ABU18309 Protein e
53	422	22.4	363	5	AB48474 Listeria
54	422	22.4	363	6	ABU32852 Protein e
55	420	22.3	362	6	ABU29950 Protein e
56	420	22.3	363	7	ADC97130 E. faeciu
57	411.5	21.9	363	6	ABU14640 Protein e
58	406.5	21.6	387	6	ADB11850 Alloioococ
59	402.5	21.4	368	4	AAU35044 Enterococ
60	401.5	21.3	363	4	AAU33370 Enterococ
61	393.5	20.9	362	5	AAU99822 Bacillus
62	370	19.7	362	5	AAU99818 Enterococ
63	368.5	19.6	359	5	AAU99819 Enterococ
64	335.5	17.8	354	6	ABU24880 Protein e
65	332	17.7	431	5	ABP91578 Herbicida
66	328.5	17.5	359	6	ABU23929 Protein e
67	315.5	16.8	384	6	ABU48604 Protein e
68	313.5	16.7	385	5	AAU99821 Rickettsi
69	310	16.5	326	2	AAU37382 Protein 1
70	307	16.3	352	7	ADD43922 Chlamydia
71	303	16.1	357	2	AAU35559 Chlamydia
72	284.5	15.1	357	2	AAU35559 Chlamydia
73	284.5	15.1	357	5	ABP90579 Chlamydia
74	284.5	15.1	357	5	ABU27032 Streptoco
75	279	14.8	353	5	AAU99820 Streptoco
76	277	14.7	353	4	AAU35971 Helicobac
77	275.5	14.6	352	4	AAU38041 Streptoco
78	275	14.6	353	6	ABU30968 Protein e
79	273.5	14.5	352	2	AAW51348 Streptoco
80	273	14.5	356	4	AAU37021 Staphyloc
81	273	14.5	356	4	AAU37260 Staphyloc
82	273	14.5	356	6	ABU42410 Protein e
83	273	14.5	356	6	ABM73029 Streptoco
84	272.5	14.5	352	4	AAU37806 Streptoco
85	272.5	14.5	352	6	ABU45942 Protein e
86	269.5	14.3	352	6	ABU01073 S. pneumo
87	263	14.0	342	6	ABU26566 Protein e
88	261	13.9	357	6	ABU42647 Protein e
89	261	13.9	366	5	ABP38353 Staphyloc
90	260	13.8	367	4	AAAG1658 S. epide
91	259.5	13.8	353	6	ABU19335 Protein e
92	250	13.3	360	5	ABP29983 Streptoco
93	250	13.3	360	6	ABU46729 Protein e
94	250	13.3	367	5	ABP25519 Streptoco
95	247	13.3	358	5	ABP25518 Streptoco
96	224	11.9	313	4	AAAG2061 S. epide
97	213.5	11.4	361	6	ABU44117 Protein e
98	160.5	8.5	186	6	ABU43476 Protein e

99 154 8.2 104 5 ABP11313 Human ORF
100 150 8.0 554 4 AAG80912 MGD synth

ALIGNMENTS

RESULT 1

AAU99816
ID AAU99816 standard; protein; 364 AA.

XX AC AAU99816;

XX DT 07-OCT-2002 (first entry)

XX DE Escherichia coli Membrane associated UDP-glycosyltransferase MurG.

XX KW MurG; membrane associated UDP-glycosyltransferase; antibiotic;
XX KW antimicrobial; modulator of glycosyltransferase activity; drug design;
XX KW UDP-glycosyltransferase; directed drug design; random drug design;
XX KW grid-based drug design.

XX OS Escherichia coli.

XX FH Key Location/Qualifiers

XX FT Domain 7..163

XX FT Region /label= N-domain

XX FT Misc-difference 14..19

XX FT /label= G-loop_1

XX FT Misc-difference 75

XX FT /note= "Member of the N-domain hydrophobic patch"

XX FT Misc-difference 79

XX FT /note= "Member of the N-domain hydrophobic patch"

XX FT Misc-difference 82

XX FT /note= "Member of the N-domain hydrophobic patch"

XX FT Misc-difference 85

XX FT /note= "Member of the N-domain hydrophobic patch"

XX FT Region 102..109

XX FT /label= G-loop_2

XX FT Misc-difference 116

XX FT /note= "Member of the N-domain hydrophobic patch"

XX FT Domain 164..340

XX FT /label= C-domain

XX FT Region 190..196

XX FT /label= G-loop_3

XX FT Domain 341..357

XX FT /label= N-domain

XX WO200190301-A2.

XX PD 29-NOV-2001.

XX PF 09-APR-2001; 2001WO-US011500.

XX PR 17-MAY-2000; 2000US-0204930P.

XX PA (UYPR-) UNIV PRINCETON.

XX PI Walker S, Ha S;

XX DR WPI; 2002-171402/22.

XX PT Novel composition comprising crystalline form of MurG protein, a membrane

XX PT -associated UDP-glycosyltransferase involved in peptidoglycan

XX PT biosynthesis, for determining ability of chemical compound to bind MurG

XX PT protein.

XX PS Claim 14; Fig 3A; 222pp; English.

XX CC The invention describes a composition comprising a membrane associated

XX CC UDP-glycosyltransferase, MurG, preferably Escherichia coli protein in

XX CC crystalline form. A model of UDP-glycosyltransferase is useful in a

XX CC computer-assisted method of structure based drug design of bioactive

CC compounds, by providing and designing a chemical compound using the
CC model. The method further comprises synthesising the chemical compound,
CC and evaluating the bioactivity of the synthesised chemical compound. The
CC bioactivity is selected from inhibiting binding of a nucleotide donor
CC compound or an acceptor compound to the MurG protein, or inhibiting
CC association of the MurG protein to a membrane. Designing the chemical
CC compound involves computational screening of one or more database of
CC chemical compounds in which the 3D structure of the compounds are known,
CC and interacting a compound identified by the screening step with the
CC model by computer. The step of designing involves directed drug design,
CC random drug design, or grid-based drug design. Designing involves
CC selecting compounds which are predicted to bind to or mimic the 3D
CC structure of the MurG protein. A modulator of glycosyltransferase is
CC useful as antibiotics or antimicrobial agents in animals, and
CC therapeutically or diagnostically in an animal. This is the amino acid
CC sequence of the Escherichia coli MurG protein crystallised in the
CC invention

XX SQ Sequence 364 AA;

Query Match 97.9%; Score 1841; DB 5; Length 364;

Best Local Similarity 98.1%; Pred. No. 4e-175;

Matches 357; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VMSSGQKELVMAGTGGHVPGLAVAHHLMAQGHVWLTADRMEDALVPKHGIEIDF 60
DB 1 VMSSGQKELVMAGTGGHVPGLAVAHHLMAQGHVWLTADRMEDALVPKHGIEIDF 60
QY 61 IRISGLRGKGIKALIAAPLRFNARQARAIKAYKPDVVLGMGVYSGPGLAASLGI 120
DB 61 IRISGLRGKGIKALIAAPLRFNARQARAIKAYKPDVVLGMGVYSGPGLAASLGI 120
QY 121 PVLVHEQNGIAGLTNKLARIATKMQAEPGAFNAEVVGNPVRTDVLALPLPQORLAGR 180
DB 121 PVLVHEQNGIAGLTNKLARIATKMQAEPGAFNAEVVGNPVRTDVLALPLPQORLAGR 180
QY 181 EGPVRLVVGSGQARILNQTMPQVAAKLGDSVITWHQSGKGSQSVQVAAEAGQPQHK 240
DB 181 EGPVRLVVGSGQARILNQTMPQVAAKLGDSVITWHQSGKGSQSVQVAAEAGQPQHK 240
QY 241 VTEPIDDMAAYAMADVVCSSGALITVSEIAAAGLPALFVPFQHKDROQYWNALPLEKAG 300
DB 241 VTEPIDDMAAYAMADVVCSSGALITVSEIAAAGLPALFVPFQHKDROQYWNALPLEKAG 300
QY 301 AAKIIEQPQLSVDVANTLACWSRETLTMAERARAASIIPATERVANEVSRVARALEHH 360
DB 301 AAKIIEQPQLSVDVANTLACWSRETLTMAERARAASIIPATERVANEVSRVARALEHH 360
QY 361 HHHH 364
DB 361 HHHH 364

RESULT 2

AAU34438

ID AAU34438 standard; protein; 355 AA.

XX AC AAU34438;

XX DT 14-FEB-2002 (first entry)

XX DE E. coli cellular proliferation protein #19.

XX KW Antisense; prokaryotic cellular proliferation protein; antibiotic;

XX KW antibacterial; drug design.

XX OS Escherichia coli.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207727P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI: 2001-611495/70.
 DR N-PSDB; AAS52297.
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX Example 3; SEQ ID NO 10031; 511pp; English.
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 355 AA;
 SQ

Query Match 95.9%; Score 1803; DB 4; Length 355;
 Best Local Similarity 99.2%; Pred. No. 2.5e-171;
 Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 MSGGCKRLMWAGTGGHVPGLAVAHLLMAQGQVRLWLTADRMEDLVPHKGIEDFI 61
 DB 1 MSGGCKRLMWAGTGGHVPGLAVAHLLMAQGQVRLWLTADRMEDLVPHKGIEDFI 60
 QY 62 RISGLRGKIKALIAAPLRIFNWQRARAIMKAYKPDVVLGMGGYVSGPGLAASLGIP 121
 DB 61 RISGLRGKIKALIAAPLRIFNWQRARAIMKAYKPDVVLGMGGYVSGPGLAASLGIP 120
 QY 122 VVLEHQNGIAGLTNKLWARIATKWMQAEPPNAEVVGNPVRTDVLALPLPQORLAGRE 181
 DB 121 VVLEHQNGIAGLTNKLWARIATKWMQAEPPNAEVVGNPVRTDVLALPLPQORLAGRE 180
 QY 182 GPVRLVVGSGQARILNQTMQVAAKLGDSVITWHQSGKGSQSQVEQAYAEAGQPQHKV 241
 DB 181 GPVRLVVGSGQARILNQTMQVAAKLGDSVITWHQSGKGSQSQVEQAYAEAGQPQHKV 240
 QY 242 TEFTDDMAAYADVVCVRSGLTIVSEIAAGLPALFVFPQHKDQQYNNALPLEKAGA 301
 DB 241 TEFTDDMAAYADVVCVRSGLTIVSEIAAGLPALFVFPQHKDQQYNNALPLEKAGA 300
 QY 302 AKIIEQPQLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVARA 356
 DB 301 AKIIEQPQLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 3

AAAG98406
 ID AAG98406 standard; protein; 355 AA.
 XX
 AC AAG98406;
 DT 21-SEP-2001 (first entry)
 XX
 DE Escherichia coli protein sequence SEQ ID NO:454.
 XX
 KW Escherichia coli; identification; proliferation; microorganism;
 KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
 KW bacterial growth inhibition.
 XX
 OS Escherichia coli.
 XX
 PN WO200148209-A2.
 XX
 PD 05-JUL-2001.
 XX
 PF 19-DEC-2000; 2000WO-US034419.
 XX
 PR 23-DEC-1999; 99US-0173005P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Forsyth RA, Ohlsen KL, Zyskind JW;
 XX WPI: 2001-457376/49.
 DR N-PSDB; AAH81462.
 XX
 PT Novel nucleic acids encoding proteins required for Escherichia coli
 PT proliferation, useful for screening for antimicrobial agents.
 XX
 PS Claim 19; Page 573; 596pp; English.
 XX
 CC The present invention describes a purified or isolated nucleic acid
 CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
 CC given in AAH81292 to AAH81294, where expression of the nucleic acid in a
 CC microorganism is capable of inhibiting proliferation of a microorganism.
 CC (I) have antibacterial and antibiotic activities, and can be used in gene
 CC therapy. Expression of (I) in a microorganism inhibits proliferation of
 CC the microorganism, and the manufactured antibiotic is useful for reducing
 CC the activity or level of a gene product required for proliferation of a
 CC microorganism in a subject, specifically humans. The nucleic acids that
 CC inhibit bacterial growth or proliferation can be used as antisense
 CC therapeutics for killing bacteria. In addition to therapeutic
 CC applications, the nucleic acid sequences complementary to sequences
 CC required for proliferation can be used as diagnostic tools. For example,
 CC nucleic acid probes complementary to proliferation-required sequences
 CC that are specific for particular species of microorganisms can be used as
 CC probes to identify particular microorganism species in clinical
 CC specimens. AAH81295 to AAH81487 encode the Escherichia coli proteins
 CC given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent
 CC oligonucleotides, which are used in the exemplification of the present
 CC invention
 XX Sequence 355 AA;
 SQ

Query Match 95.9%; Score 1803; DB 4; Length 355;
 Best Local Similarity 99.2%; Pred. No. 2.5e-171;
 Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 MSGGCKRLMWAGTGGHVPGLAVAHLLMAQGQVRLWLTADRMEDLVPHKGIEDFI 61
 DB 1 MSGGCKRLMWAGTGGHVPGLAVAHLLMAQGQVRLWLTADRMEDLVPHKGIEDFI 60
 QY 62 RISGLRGKIKALIAAPLRIFNWQRARAIMKAYKPDVVLGMGGYVSGPGLAASLGIP 121
 DB 61 RISGLRGKIKALIAAPLRIFNWQRARAIMKAYKPDVVLGMGGYVSGPGLAASLGIP 120
 QY 122 VVLEHQNGIAGLTNKLWARIATKWMQAEPPNAEVVGNPVRTDVLALPLPQORLAGRE 181
 DB 121 VVLEHQNGIAGLTNKLWARIATKWMQAEPPNAEVVGNPVRTDVLALPLPQORLAGRE 180

QY 182 GPRVVLVVGSGQARILNQTMPQVAAKLGDSVITWHQSGKSGSQSQVEQAYAEAGQPQHKV 241
DB 181 GPRVVLVVGSGQARILNQTMPQVAAKLGDSVITWHQSGKSGSQSQVEQAYAEAGQPQHKV 240
QY 242 TEFIDDDMAAYANADVVCVSGALTVSEIAAAGLPALFVFPQHKDRQQYWNALPLEKAGA 301
DB 241 TEFIDDDMAAYANADVVCVSGALTVSEIAAAGLPALFVFPQHKDRQQYWNALPLEKAGA 300
QY 302 AKIIIEQQLSVDVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVARA 356
DB 301 AKIIIEQQLSVDVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 4
ABU28491
ID ABU28491 standard; protein; 355 AA.
XX AC ABU28491;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #14018.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Escherichia coli.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 08-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX DR N-PSDB; ACA32361.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 56415; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 5213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 355 AA;

Query Match 95.9%; Score 1803; DB 6; Length 355;
Best Local Similarity 99.2%; Pred. No. 2.5e-171;
Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 MSGGKRLMVMAGGTGGHVFPGLAVAHHLMAQGVQVRLGTADRMEADLVPKHGIEIDFI 61
DB 1 MSGGKRLMVMAGGTGGHVFPGLAVAHHLMAQGVQVRLGTADRMEADLVPKHGIEIDFI 60
QY 62 RISGLRGKIGIKALIAAPLIRIFNWRQARAIWKAYKPDVILGMGGYVSGPGLAAMSIGIP 121
DB 61 RISGLRGKIGIKALIAAPLIRIFNWRQARAIWKAYKPDVILGMGGYVSGPGLAAMSIGIP 120
QY 122 VTLHEQNGIAGLTNKLARIATKYNQABPGAFPAEYVGVNPNFTDYLALPLPQORLAGRE 181
DB 121 VTLHEQNGIAGLTNKLARIATKYNQABPGAFPAEYVGVNPNFTDYLALPLPQORLAGRE 180
QY 182 GPRVVLVVGSGQARILNQTMPQVAAKLGDSVITWHQSGKSGSQSQVEQAYAEAGQPQHKV 241
DB 181 GPRVVLVVGSGQARILNQTMPQVAAKLGDSVITWHQSGKSGSQSQVEQAYAEAGQPQHKV 240
QY 242 TEFIDDDMAAYANADVVCVSGALTVSEIAAAGLPALFVFPQHKDRQQYWNALPLEKAGA 301
DB 241 TEFIDDDMAAYANADVVCVSGALTVSEIAAAGLPALFVFPQHKDRQQYWNALPLEKAGA 300
QY 302 AKIIIEQQLSVDVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVARA 356
DB 301 AKIIIEQQLSVDVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 5

ABU47265
ID ABU47265 standard; protein; 355 AA.
XX AC ABU47265;
XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #32792.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Salmomella typhimurium.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 DR MPI; 2003-029926/02.
 DR N-FSDB; ACAS1135.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 XX
 PS Claim 25; SEQ ID NO 75189; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 355 AA;

Query Match 89.7%; Score 1687; DB 6; Length 355;
 Best Local Similarity 91.2%; Pred. No. 9.9e-160;
 Matches 323; Conservative 18; Mismatches 13; Indels 0; Gaps 0;
 QY 2 MSGGKELVMAGTGGHVPFGLAVAHHLMAQGWVRLGTADRMEADLVPHGIEIDFI 61
 DB 1 MSGQPRRLVMAGTGGHVPFGLAVAHHLMAQGWVRLGTADRMEADLVPHGIDIDFI 60
 QY 62 RISLRGKGIKALIAPLRIFNARQARAIMKAYKPDVVLGMGGYVSGPGLAANSLGIP 121
 DB 61 RISLRGKGVKALLAALRIFNARQARAIMKREFDVLGMGGYVSGPGLAANSLGIP 120
 QY 122 VULHEQNGIAGLTNKLARIATKMQAEFGAPFNABVGNPVRTDVLPLPQOBLAGE 181
 DB 121 VULHEQNGIAGLTNQLAKIATTWMOAFPCAPFNABVGNPVRTDVLPLPQOVLGRD 180
 QY 182 GPRVRLVVGSGQARTLNQTPQVAAKLGDVLIWHOSGKGSQSVQEOAYAGAGOPQHKV 241
 DB 181 GPRVRLVVGSGQARVLTNTPQVAAKLGDVLIWHOSGKQALTVQAYAGAGOPQHKV 240
 QY 242 TEFIDNMAAYAWADVVCVRSAGLTVSEIAAGLPAIFVFPQHKDRQQVWNPALPLEKAGA 301
 DB 241 TEFIDNMAAYAWADVVCVRSAGLTVSEIAAGLPAIFVFPQHKDRQQVWNPALPLENAGA 300
 QY 302 AKTIEQPOLSVDAVANTLAGWRETLTWTAEARARASIPDATERVANEVSRRAR 355
 DB 301 AKTIEQOFTVEAVADTLAGWREALLTWAERARAVSIPDATERVANEVSRRAR 354

RESULT 6

AAU38485
 ID AAU38485 standard; protein; 355 AA.
 XX
 AC AAU38485;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Salmonella typhi cellular proliferation protein #376.
 XX
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.
 XX
 OS Salmonella typhi.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US009180.
 XX
 PR 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207272P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KI, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR MPI; 2001-611495/70.
 DR N-FSDB; AAS56344.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX
 PS Example 3; SEQ ID NO 14078; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
 CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
 CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 355 AA;
 Query Match 89.2%; Score 1677; DB 4; Length 355;
 Best Local Similarity 90.7%; Pred. No. 9.9e-159;
 Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;
 QY 2 MSGGKELVMAGTGGHVPFGLAVAHHLMAQGWVRLGTADRMEADLVPHGIEIDFI 61
 DB 1 MSGQPRRLVMAGTGGHVPFGLAVAHHLMAQGWVRLGTADRMEADLVPHGIDIDFI 60
 QY 62 RISLRGKGIKALIAPLRIFNARQARAIMKAYKPDVVLGMGGYVSGPGLAANSLGIP 121

Db 61 RISLGRGKGVKALLAALPLRIFNAWQARAIMKRFDPDVLVGGVSGPGLAANSLGIP 120
QY 122 VVLHEONGIAGLTNKLARIATKVMQAEFGAPFNAEVGNVVRTDVLALPIPOORLAGRE 181
Db 121 VVLHEONGIAGLTNQLAKIATVMQAEFGAPFNAEVGNVVRTDVLALPIPOORLAGRD 180
QY 182 GPRVLVVGSGQARILNQTMPQVAAKLGDSVIIHQSGKSGQSQVQAYAEAGOPQHKV 241
Db 181 GPRVLVVGSGQARVLTNQTMPQVAAKLGDSVIIHQSGKSGQALTVQAYAGAGOPQHKV 240
QY 242 TEFIDMAAAYAWADVVCVRSGLATVSEIAAAGLPALFVFPFHQKDRQYWNALPLEKAGA 301
Db 241 TEFIDMAAAYAWADVVCVRSGLATVSEIAAAGLPALFVFPFHQKDRQYWNALPLENAGA 300
QY 302 AKIEPQLSDVAVANTLAGSRETLTWAERARAASIPDATERVANEVSRVAR 355
Db 301 AKIEPQPTVEAVADTLAGSREALLTWAERARAVSIPDATERVASEVSRVAR 354

RESULT 7
ABU48097
ID ABU48097 standard; protein; 355 AA.
XX AC ABU48097;
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #33624.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Salmomella typhi.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029826/02.
DR N-PSDB; ACA51967.

XX New Antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 76021; 1765pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 355 AA;

Query Match 89.2%; Score 1677; DB 6; Length 355;
Best Local Similarity 90.7%; Pred. No. 9.9e-159;
Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;
QY 2 MSGQKRLVMAGTGGHVPFGLAVAHLMQAGQVRLGTADRMADLVPKHGIDIFI 61
Db 1 MSGQKRLVMAGTGGHVPFGLAVAHLMQAGQVRLGTADRMADLVPKHGIDIFI 60
QY 62 RISLGRGKGVKALLAALPLRIFNAWQARAIMKRFDPDVLVGGVSGPGLAANSLGIP 121
Db 61 RISLGRGKGVKALLAALPLRIFNAWQARAIMKRFDPDVLVGGVSGPGLAANSLGIP 120
QY 122 VVLHEONGIAGLTNKLARIATKVMQAEFGAPFNAEVGNVVRTDVLALPIPOORLAGRE 181
Db 121 VVLHEONGIAGLTNQLAKIATVMQAEFGAPFNAEVGNVVRTDVLALPIPOORLAGRD 180
QY 182 GPRVLVVGSGQARILNQTMPQVAAKLGDSVIIHQSGKSGQSQVQAYAEAGOPQHKV 241
Db 181 GPRVLVVGSGQARVLTNQTMPQVAAKLGDSVIIHQSGKSGQALTVQAYAGAGOPQHKV 240
QY 242 TEFIDMAAAYAWADVVCVRSGLATVSEIAAAGLPALFVFPFHQKDRQYWNALPLEKAGA 301
Db 241 TEFIDMAAAYAWADVVCVRSGLATVSEIAAAGLPALFVFPFHQKDRQYWNALPLENAGA 300
QY 302 AKIEPQLSDVAVANTLAGSRETLTWAERARAASIPDATERVANEVSRVAR 355
Db 301 AKIEPQPTVEAVADTLAGSREALLTWAERARAVSIPDATERVASEVSRVAR 354

RESULT 8
ABU31657
ID ABU31657 standard; protein; 348 AA.
XX AC ABU31657;
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #17184.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Klebsiella pneumoniae.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-029926/02.
XX N-PSDB; ACA35527.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 59581; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 348 AA;
XX
XX Query Match 85.3%; Score 1604; DB 6; Length 348;
XX Best Local Similarity 88.8%; Pred. No. 1.9e-151;
XX Matches 308; Conservative 19; Mismatches 20; Indels 0; Gaps 0;
XX
XX 10 MNVAGTGGHVFVGLAVAHLMQAGQVRVLGTDADMEADLVPKHGIEIDFIRISGLRK 69
XX 1 MNVAGTGGHVFVGLAVAHLMQAGQVRVLGTDADMEADLVPKHGIEIDFIRISGLRK 60
XX
XX 70 GIKALINAPRIINAWQARAIKAYKPDVVLGNGGVSGPGGLAWSIGIPVWLHEQNG 129
XX 61 GIKAQLLAPRIINAWQARAIKAYKPDVVLGNGGVSGPGGLAWSIGIPVWLHEQNG 120
XX 130 IAGLTNKLARIATKVMQABPFGAFFNAEVVGNVPRVTDVIALPLPQORLAGRGGPVRVLVV 189
XX 121 IAGLTNKLARIATKVMQABPFGAFFNAEVVGNVPRVTDVIALPLPQORLAGRGGPVRVLVV 180
XX 190 GSGQGARILNQTMPQAAKLGDSVVIWHQSGKGSQSVQAYAEACQPHKVTETIDMA 249
XX 181 GSGQGARVLNQTMPQAAKLGATVTIWHQSGKGSQSVQAYAEACQPHKVTETIDMA 240
XX 250 AAYAWADVVCRSGLTIVSEIAAGLPALFVPFHQKDRQYWNALPLEKAGAAKILEQPE 309
XX

Db 241 AAYAWADVVCRSGLTIVSEIAAGLPALFVPFHQKDRQYWNALPLEKAGAAKILEQPE 300
QY 310 LSVDAVANTLACMSRETLITWARRARASIPDATERVANEYSRVARA 356
Db 301 FTVEAVASTLASWDRETLIDWAERGASIPDATERVAAEVSVALA 347
RESULT 9
ABU45116
ID ABU45116 standard; protein; 348 AA.
XX ABU45116;
XX DT 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #30643.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Salmonella paratyphi.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US0009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-029926/02.
XX N-PSDB; ACA48986.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 73040; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 348 AA;
XX
XX Query Match 85.3%; Score 1604; DB 6; Length 348;
XX Best Local Similarity 88.8%; Pred. No. 1.9e-151;
XX Matches 308; Conservative 19; Mismatches 20; Indels 0; Gaps 0;
XX
XX 10 MNVAGTGGHVFVGLAVAHLMQAGQVRVLGTDADMEADLVPKHGIEIDFIRISGLRK 69
XX 1 MNVAGTGGHVFVGLAVAHLMQAGQVRVLGTDADMEADLVPKHGIEIDFIRISGLRK 60
XX
XX 70 GIKALINAPRIINAWQARAIKAYKPDVVLGNGGVSGPGGLAWSIGIPVWLHEQNG 129
XX 61 GIKAQLLAPRIINAWQARAIKAYKPDVVLGNGGVSGPGGLAWSIGIPVWLHEQNG 120
XX 130 IAGLTNKLARIATKVMQABPFGAFFNAEVVGNVPRVTDVIALPLPQORLAGRGGPVRVLVV 189
XX 121 IAGLTNKLARIATKVMQABPFGAFFNAEVVGNVPRVTDVIALPLPQORLAGRGGPVRVLVV 180
XX 190 GSGQGARILNQTMPQAAKLGDSVVIWHQSGKGSQSVQAYAEACQPHKVTETIDMA 249
XX 181 GSGQGARVLNQTMPQAAKLGATVTIWHQSGKGSQSVQAYAEACQPHKVTETIDMA 240
XX 250 AAYAWADVVCRSGLTIVSEIAAGLPALFVPFHQKDRQYWNALPLEKAGAAKILEQPE 309
XX

CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 348 AA;

Query Match 82.4%; Score 1549.5; DB 6; Length 348;
Best Local Similarity 86.3%; Pred. No. 5.5e-146;
Matches 302; Conservative 16; Mismatches 30; Indels 1; Gaps 1;
QY 2 MSGQKRLVMAGTGGHVPFGLAVAHLMAGQVVRWLTADRMEDLVPKHGIEIDFI 61
Db 1 MSGQKRLVMAGTGGHVPFGLAVAHLMAGQVVRWLTADRMEDLVPKHGIEIDFI 60
QY 62 RISGLRGKIKALIAAPLRFINAWROARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 121
Db 61 RISGLRGKIKALIAAPLRFINAWROARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 120
QY 122 VVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAEVGNPVRTDVLALPLPQORLAGRE 181
Db 121 VVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAEVGNPVRTDVLALPLPQORLAGRD 180
QY 182 GPRVRLVVGSGQARILNCTMPQVAAKLGDVLIHQSGKSGQSQVQAYABAGOPQHKV 241
Db 181 GPRVRLVVGSGQARILNCTMPQVAAKLGDVLIHQSGKSGQALTVQAYAGTGPQHKV 240
QY 242 TEFDIDMAAYAWADVVCESGALTVSEIAAGLPALFVFFQHKDRQQYWNALPLEKAGA 301
Db 241 TEFDIDMAAYAWADVVCESGALTVSEIAXPLGPAIFVFFQHKDRQQYWNALPLENAGA 300
QY 302 AKTIEQPQLSVDAVANTLAGWSRETLITMAERARAASIPDATERVANEV 350
Db 301 AKTIEQPQFTVDSPT-GGVAAELTTWAEVPAVSPDATERVASEV 348

RESULT 10
ABU50002
ID ABU50002 standard; protein; 356 AA.
AC ABU50002;
XT 19-JUN-2003 (first entry)
DE Protein encoded by Prokaryotic essential gene #35529.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Yersinia pestis.
OS WO200277183-A2.
PN 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-0299926/02.
DR N-PSDB; ACA53872.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT

PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 77926; 1766pp; English.
PS The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 356 AA;

Query Match 78.3%; Score 1471; DB 6; Length 356;
Best Local Similarity 78.3%; Pred. No. 4e-138;
Matches 278; Conservative 36; Mismatches 41; Indels 0; Gaps 0;
QY 2 MSGQKRLVMAGTGGHVPFGLAVAHLMAGQVVRWLTADRMEDLVPKHGIEIDFI 61
Db 1 MSGQKRLVMAGTGGHVPFGLAVAHLMAGQVVRWLTADRMEDLVPKHGIEIDFI 60
QY 62 RISGLRGKIKALIAAPLRFINAWROARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 121
Db 61 RISGLRGKIKALIAAPLRFINAWROARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 120
QY 122 VVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAEVGNPVRTDVLALPLPQORLAGRE 181
Db 121 VVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAEVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPRVRLVVGSGQARILNCTMPQVAAKLGDVLIHQSGKSGQSQVQAYABAGOPQHKV 241
Db 181 GPRVRLVVGSGQARILNCTMPQVAAKLGDVLIHQSGKSGQSQVQAYABAGOPQHKV 240
QY 242 TEFDIDMAAYAWADVVCESGALTVSEIAAGLPALFVFFQHKDRQQYWNALPLEKAGA 301
Db 241 TEFDIDMAAYAWADVVCESGALTVSEIAXPLGPAIFVFFQHKDRQQYWNALPLEKAGA 300
QY 302 AKTIEQPQLSVDAVANTLAGWSRETLITMAERARAASIPDATERVANEVSRVRA 356
Db 301 AKTIEQPQFTATSVSLLASWDRATLLSMAERARSVAIPDATERVAEEVVAASKS 355
RESULT 11
ABM68867
ID ABM68867 standard; protein; 356 AA.
XX ABM68867;
XX

DT 20-NOV-2003 (first entry)
XX Photorhabdus luminescens protein sequence #1964.
DE Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX Photorhabdus luminescens.
OS WO200294867-A2.
XX 28-NOV-2002.
XX 07-FEB-2002; 2002WO-IB003040.
XX 07-FEB-2001; 2001FR-00001659.
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX WPI; 2003-148459/14.
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX Claim 2; SEQ ID NO 1964; 1205pp; French.
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterials useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically (to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This
XX sequence represents one of the isolated P. luminescens proteins.
SQ Sequence 356 AA;
Query March 76.1%; Score 1431; DB 6; Length 356;
Best Local Similarity 76.3%; Pred. No. 4e-134;
Matches 264; Conservative 44; Mismatches 38; Indels 0; Gaps 0;
QY 2 MSGQKRLMYAGTGTHVPPGLAVHLLMAQGVQVRLGTADRMADLVPRKHGIEDFI 61
DB 1 MSGKTRLMYAGTGTHVPPGLAVHLLMAQGVQVRLGTADRMADLVPRKHGIEDFI 60
QY 62 RISLRGKGKIALIAPLRIFNARQARATMKAYKEDVVLGMGYVSGGGLAANSLGIP 121
DB 61 QISLRGKGKIALIAPLRIFNARQARATMKAYKEDVVLGMGYVSGGGLAANSLGIP 120
QY 122 VVLHEQNGIAGLNKMLARIATKVMQAEPCGAFNAEVGNPVRTDLVLPQRLAGRE 181
DB 121 VVLHEQNGIAGLNKMLARIATKVMQAEPCGAFNAEVGNPVRTDLVLPQRLAGRE 180
QY 182 GPRVRLVVGSGQCARILNQMPPEIARMDGDKITLWHQTKGAKESVQYAYNSVKEHKI 240

DB 181 GPRVRLVVGSGQCARILNQMPPEIARMDGDKITLWHQTKGAKESVQYAYNSVKEHKI 240
QY 242 TEFIDMAAAYAWADVVCVRSALTVSIIAAGLPALFVFPQHKDRQYVWALPLEKAGA 301
DB 241 TEFIDMAAAYAWADVVCVRSALTVSIIAAGLPALFVFPQHKDRQYVWALPLEKAGA 300
QY 302 AKIIEQPOLSDVAVANTLAGNSRETLTMAERARASIPDATERVA 347
DB 301 AKIIEQPOLSDVAVANTLAGNSRETLTMAERARASIPDATERVA 346
RESULT 12
ABU40887
ID ABU40887 standard; protein; 360 AA.
XX AC ABU40887;
XX 19-JUN-2003 (first entry)
DT Protein encoded by Prokaryotic essential gene #26414.
DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Proteus sp.
XX WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00915242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA44757.
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 68811; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 71.5%; Score 1345.5; DB 6; Length 360;
Best Local Similarity 72.1%; Pred. No. 1.5e-125;
Matches 259; Conservative 40; Mismatches 55; Indels 5; Gaps 1;

QY 2 MSQGGKRLVMAGTGGHVPFGLAVAHILMAQGVRLGTADRMEDLVPKHGIEIDFI 61
DB 1 MSERKRLVMAGTGGHVPFGLAVAHILMAQGVRLGTADRMEDLVPKHGIEIYI 60
QY 62 RISGLRGKIGKALIAAPLIRIFNAWRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 121
DB 61 RISGLRGKGVKALIAAPIRIKAFQARRIMKEYQFDVAVLGMGGYVSGPGLAASCGIP 120
QY 122 VILHEONGIAGLTNKLARIATKVMQAEPCGAPNAEVVGNFVRTDVLALPLFOORLAGRE 181
DB 121 VILHEONGIAGLTNKLARIATKVMQAEPCGAPNAEVVGNFVRTDVLALPLFOORLAGRE 180
QY 182 GPRVRLVVGSGQARILNQTMPQVAAKLGDVSLIWHQSGKGSQSQVEQAY-----AAGQ 236
DB 181 GAVRLVVGSGQARILNQTMPQVAAKLGDVSLIWHQSGKGSQSQVEQAY-----AAGQ 240
QY 237 PQHKVTEFIDDMAAAYAVADVVYCRSGALTVSIIAAGLPALFVPPQHKDROQYWNALPL 296
DB 241 SEYKVTIEFIDIAQAYQADVVYCRSGALTVSIIAAGLPALFVPPQHKDROQYWNALPL 300
QY 297 EKAGAKIIEPOI.SYDVANTLAGWSRETLTMAERARAAASTPDATERVANEVSRRAR 355
DB 301 ENAGAKIIEQNDLTFEATADLTLENDRHQLMLMAEKAGSVAITDATERVANVITEVAK 359

RESULT 13
ABU27914
ID ABU27914 standard; protein; 281 AA.
XX
AC ABU27914;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #13441.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Enterobacter cloacae.
XX
FN W0200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 28-OCT-2001; 2001US-0342923P.
XX
PR 08-FEB-2002; 2002US-00072851.
XX
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX

DR MPI; 2003-029926/02.
XX N-PSDB; ACA31784.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 55938; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a gene in an operon that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 281 AA;

Query Match 70.3%; Score 1323; DB 6; Length 281;
Best Local Similarity 91.3%; Pred. No. 1.8e-123;
Matches 253; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 5 QGKELMWAGTGGHVPFGLAVAHILMAQGVRLGTADRMEDLVPKHGIEIDFI 64
DB 3 QPKELMWAGTGGHVPFGLAVAHILMAQGVRLGTADRMEDLVPKHGIEIDFI 62
QY 65 GLRGKIGKALIAAPLIRIFNAWRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIPVL 124
DB 63 GLRGKIGKALIAAPLIRIFNAWRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIPVL 122
QY 125 HEONGIAGLTNKLARIATKVMQAEPCGAPNAEVVGNFVRTDVLALPLFOORLAGRGPV 184
DB 123 HEONGIAGLTNKLARIATKVMQAEPCGAPNAEVVGNFVRTDVLALPLFOORLAGRGPV 182
QY 185 RVLVVGSGQARILNQTMPQVAAKLGDVSLIWHQSGKGSQSQVEQAYAEAGQPHKYTEF 244
DB 183 RVLVVGSGQARILNQTMPQVAAKLGDVSLIWHQSGKGSQSQVEQAYAEAGQPHKYTEF 242
QY 245 IDDMAAYAVADVVYCRSGALTVSEIAAAGLPALFVP 281
DB 243 IDDMAAYAVADVVYCRSGALTVSEIAAAGLPALFVP 279

RESULT 14
ABU49532
ID ABU49532 standard; protein; 354 AA.
XX
XX ABU49532;
AC

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 354 AA;

Query Match 59.7%; Score 1123; DB 6; Length 354;
 Best Local Similarity 63.1%; Pred. No. 2.6e-103;
 Matches 222; Conservative 46; Mismatches 76; Indels 8; Gaps 2;
 QY 2 MSQGGKLVMAAGTGGHVPFGLVAHLMAGQVRLWGLTADRMEDLVPHKGIIDIFI 61
 DB 1 MSEQKRLVMAAGTGGHVPFGLVAHLMAGQVRLWGLTADRMEDLVPHKGIIDIFI 60
 QY 62 RISLGRKGKIKALIAAPLRFNMRQARAIMKAYPDVVLGMYGVSPPGGLAANSLGIP 121
 DB 61 QISLGRKGKIKALISAPPSIFRAILQARIIKAYOPHVLGMYGVSPPGGLAANSLGIP 120
 QY 122 VULHEONGIAGLTKMLARIATKMOAEPGAPNAEVGNVPTDVLALPLPQORLAGRE 181
 DB 121 VILHEQNAVAGLNSWLAKIARRVLQAFPTAFPAEAVGNVPTDVLALPLPQORLAGRE 180
 QY 182 GPRVRLVWGSQGARILNQTMPQVAAKLGDSVLIHQSGKGSQSQSVQAYAEAGQPH-- 239
 DB 181 GKRLVWGSQGARILNQTLPNVAAQLSDKLEVRHQVGQAVEQVTTLY----PEHAS 235
 QY 240 -KYTEFIDMAAYADVVVCRSGALTYSSEIAAGLPALFPFQHKDQOYWNALPLEK 298
 DB 236 VTITTEFIDNADAYADIVICRSGALTYSSEIAAGLPALFPFQHKDQOYWNALPLEK 295
 QY 299 AGAAKIEOPQSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEV 350
 DB 296 AGAATIPQAEULTAEKLVSVLTQFDRETLQQAIAKAKAMATPLAQRVAEVI 347

RESULT 16
 AAU35596
 ID AAU35596 standard; protein; 351 AA.
 XX
 AC AAU35596;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Haemophilus influenzae cellular proliferation protein #237.
 XX
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.
 XX
 OS Haemophilus influenzae.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.
 PF
 XX 21-MAR-2000; 2000US-0191078P.
 XX 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207727P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
 PI Yamamoto RT, Xu RH;
 PI N-PSDB; AAS53455.
 DR WPI; 2001-611495/70.
 XX
 DR New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX
 PS Example 3; SEQ ID NO 11189; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
 CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
 CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 351 AA;

Query Match 56.6%; Score 1064.5; DB 4; Length 351;
 Best Local Similarity 61.5%; Pred. No. 1.8e-97;
 Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;
 QY 5 QGKELVMAGTGGHVPFGLVAHLMAGQVRLWGLTADRMEDLVPHKGIIDIFI 64
 DB 2 KNKLLVMAGTGGHVPFGLVAHLMAGQVRLWGLTADRMEDLVPHKGIIDIFI 61
 QY 65 GLRGKIKALIAAPLRFNMRQARAIMKAYPDVVLGMYGVSPPGGLAANSLGIPVL 124
 DB 62 GLRGKIKALINAPFAFRAVLQAKKIQEKEKDAVLGMYGVSPPGGLAANSLGIPVL 121
 QY 125 HEONGIAGLTKMLARIATKMOAEPGAPNAEVGNVPTDVLALPLPQORLAGREPV 184
 DB 122 HEQNAIAGLTKMLGIATCVLQAFPTAFPAEAVGNVPTDVLALPLPQORLAGREPV 181
 QY 185 RVLVVGSGGARILNQTMPQVAAKLGDSVLIHQSGKGSQSQSVQAYAEAGQPHKVT 244
 DB 182 RVLVVGSGGARILNQTMPQVAAKLGDSVLIHQSGKGSQSQSVQAYAEAGQPHKVT 240
 QY 245 IDDMAAYADVVVCRSGALTYSSEIAAGLPALFPFQHKDQOYWNALPLEKAGAKI 304
 DB 241 IDNMAAYADVVVCRSGALTYSSEIAAGLPALFPFQHKDQOYWNALPLEKAGAKI 300
 QY 305 IEQPLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSR 352
 DB 301 IEQADLTPEILLVNLKNTRENLLQALKATNSMFPNAAQRVAEVIKQ 348

RESULT 17

ABU30498
ID ABU30498 standard; protein; 351 AA.XX AC ABU30498;
XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #16025.

XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX KW Haemophilus influenzae.
XX OS WO200277183-A2.
XX FN 03-OCT-2002.
XX PD 21-MAR-2002; 2002WO-US009107.
XX PF 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX XX (ELIT-) ELITRA PHARM INC.XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX XXDR WPI; 2003-029926/02.
DR N-PSDB; ACA34368.XX XX New antisense nucleic acids, useful for identifying proteins or screening
XX XX for homologous nucleic acids required for cellular proliferation to
XX XX isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 58422; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: the sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 351 AA;

Query Match 56.6%; Score 1064.5; DB 6; Length 351;
Best Local Similarity 61.5%; Pred. No. 1.8e-97;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;

QY	5	CGKRLMVMAGGTGGHVFPGGLAVAHHLMAQGWQVWLGTADTADMEADLVPKHGIEIDFIRIS	64
DB	2	KNKLLVMAGGTGGHVFPGGLAVAHHLMAQGWQVWLGTADTADMEADLVPKHGIEIDFIRIS	61
QY	65	GLRGKGIKALIAAPLIRIFNAWRQARAKMAYKPDVILGNGGVSGPGGLAASLIGIPVL	124
DB	62	GLRGKGIKALLNAPFAIPRAVLQAKTIQEEKPDVILGNGGVSGPGGLAASLIGIPVL	121
QY	125	HEONGIAGLTNKLARIATKVMQAEPCGAFNAEVGVNVRDVLALPLPOQLAGREGPV	184
DB	122	HEONAIAGLTNKLKLIATCVLQAFPAFPAHVAEVGVNVRDVLALPLPOQLAGREGPV	181
QY	185	RVLVVGSGQARILNQTMPQVAAKLGDSVVIWHOSGKGSQOSVQAYABACQPQHKVTEF	244
DB	182	RVLVVGSGQARVNLHTLPKVAQLADKLEFRHVGKGAVEEVSQLYGE-NLEQVKITEF	240
QY	245	IDDMAAYAWADVVVCRSGALTVSIEIAAGLPALFVPFHQKDRQOYKVALPLEKAGAAKI	304
DB	241	IDDMAAYAWADVVVCRSGALTVCISIAAVGAAAFVFPFHQKDRQOYLNKYLSDVGAAKI	300
QY	305	IEQPOLSVDAVANTILAGWSRETLITMAERARAASIPDATERVANEVSR	352
DB	301	IEQADITPEILLVNYLKNLTRENLQWALKAKTMSMENAQAQVAEVIKQ	348

RESULT 18

AAU99817
ID AAU99817 standard; protein; 351 AA.

XX AC AAU99817;

XX DT 07-OCT-2002 (first entry)

XX DE Haemophilus influenzae membrane associated UDP-glycosyltransferase.

XX KW MurG; membrane associated UDP-glycosyltransferase; antibiotic;
XX antimicrobial; modulator of glycosyltransferase activity; drug design;
XX KW UDP-glycosyltransferase; directed drug design; random drug design;
XX grid-based drug design.

XX OS Haemophilus influenzae.

XX PH Key Location/Qualifiers

XX FT Region 11..16 /label= G-loop_1

XX FT Region 99..106 /label= G-loop_2

XX FT Region 187..193 /label= G-loop_3

XX XX WO200190301-A2.

XX XX 29-NOV-2001.

XX XX 09-APR-2001; 2001WO-US011500.

XX XX 17-MAY-2000; 2000US-0204930P.

XX XX (UYPR-) UNIV PRINCETON.

XX XX Walker S, Ha S;

XX XX WPI; 2002-171402/22.

XX XX Novel composition comprising crystalline form of MurG protein, a membrane

XX FT -associated UDP-glycosyltransferase involved in peptidoglycan

XX FT biosynthesis, for determining ability of chemical compound to bind MurG

[illegible]

RESULT 20

ABU21174
ID ABU21174 standard; protein; 372 AA.

Matches 178; Conservative 60; Mismatches 113; Indels 6; Gaps 3;

or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/bct/sequences](http://wipo.int/pub/published/bct/sequences)

Query Match	45.1%;	Score	849;	DB	6;	Length	372;
Best Local Similarity	49.94;	Pred. No.	7.2e-16;				
Matches	178;	Conservative	60;	Mismatches	113;	Indels	6;
Gaps							3
QY	7	KELVMVAGGTGGHYFPGPLVAHHLMAQGWVRVLGTADRMEADLVKPKHGIEIDFINISGL	66				
Db	7	RTLVMVAGGTGGHYFPGPLVAHLMQAWGKVMVLGNPAGMEATLVKPKHGIEIVYRFGSL	66				
QY	67	RKGKGLIALIAPLRIFNWQCARALMKAYKPDVVLGMGGVVGSPGGAAWSLIGIPVVLHE	126				
Db	67	RKGKGLTKIMPLMLNLACQTQSUSVLVRKVPDVLGMGGYITFPAGLMTALSRLPLVHE	126				
QY	127	QNGIAGLTNKKWLARIATKVMQABPGAFENNAEVGNVPRTDVLALPIPOORLAGREGVPRV	186				
Db	127	QNSIAGLANKVLAKVAKRVLVAFENALPHGEWGNPIRAELAGAIAPKARYAQRSGLPV	186				
QY	187	LVVGGSGQARLTINTQFQVAAGK--GDSVTIWHQSGKGSQOSVEQAYAEAG--QPQHKV	241				
Db	187	LVVGGSLGAAALNEVPFRAVALLAPNERPRIVHOAGAKHIEALRENYAAGLQAGADVSL	246				
QY	242	TEFTDDMAAYANADVVVCSGALTYSSEIAAGLPAIFVPFQHK--DROQYWNALPLEKAG	300				
Db	247	VFFIDDMTSAVANADLVLCISGAMTYSSEISAVGVAALFVPFPFVAVDDHQTWNAAF	306				
QY	301	AAKIEIQPLSDVAVANTLAGWSRETIITWAERARAASIPDATERVANEVSRVAARAL	357				
Db	307	AALVVOORDLSAETLADWLSORTRETLAEWAERSRLAKPDTEQVQAOATCATVAGSI	363				

Search completed: June 7, 2004, 07:13:16
Job time : 65 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 07:12:05 ; Search time 22 Seconds
(without alignments)
854.175 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1891

Sequence: 1 MMSGGKRLMAGTGGHV.....RVANEVSRALEHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database :

Issued Patents_RA*
1: /cgn2_6/protdata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/protdata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/protdata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/protdata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/protdata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/protdata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1803	95.9	355	3	US-08-984-618-14
2	1623	86.6	370	4	US-09-489-039A-12435
3	1345.5	71.5	360	4	US-09-543-681A-7067
4	1064.5	56.6	351	3	US-08-984-618-15
5	715.5	38.0	390	4	US-09-252-991A-24124
6	701.5	37.3	397	4	US-09-543-236-2678
7	680.5	36.2	366	4	US-09-328-332-5036
8	422	22.4	363	3	US-08-984-618-13
9	420	22.3	363	4	US-09-107-532A-6757
10	411.5	21.9	370	4	US-09-134-000C-4145
11	406.5	21.6	363	3	US-08-984-618-16
12	284.5	15.1	357	4	US-09-198-452A-977
13	273.5	14.5	352	2	US-08-751-474-2
14	261	13.9	366	4	US-09-134-001C-3198
15	131	7.0	525	3	US-08-984-618-17
16	127	6.8	382	3	US-08-984-618-12
17	122.5	6.5	572	4	US-09-252-991A-26203
18	117	6.2	463	4	US-08-311-731A-7
19	115.5	6.1	390	3	US-09-036-987A-8
20	115.5	6.1	390	4	US-09-370-700-8
21	115.5	6.1	390	4	US-09-603-207-8
22	114.5	6.1	537	4	US-09-653-270A-11
23	114.5	6.1	537	4	US-09-651-941-11
24	114.5	6.1	537	4	US-09-955-597-11
25	108.5	5.8	372	1	US-08-196-218-33
26	108.5	5.8	372	1	US-08-681-953-33
27	108.5	5.8	559	4	US-08-506-296B-66

28	108	5.7	743	4	US-09-252-991A-26512	Sequence 26512, A
29	106	5.6	452	4	US-08-311-731A-6	Sequence 6, Appli
30	106	5.6	1271	1	US-08-095-734-2	Sequence 2, Appli
31	106	5.6	1271	2	US-08-444-623-2	Sequence 2, Appli
32	106	5.6	1271	3	US-08-471-869-2	Sequence 2, Appli
33	106	5.6	1271	4	US-09-342-583-2	Sequence 2, Appli
34	106	5.6	1271	5	PCT-US94-08267-2	Sequence 2, Appli
35	104.5	5.6	738	4	US-09-252-991A-31185	Sequence 31185, A
36	103	5.5	746	4	US-09-252-991A-18587	Sequence 18587, A
37	103	5.5	749	4	US-09-252-991A-16977	Sequence 16977, A
38	102	5.4	1650	4	US-09-252-991A-21798	Sequence 21798, A
39	101.5	5.4	945	4	US-09-252-991A-30699	Sequence 30699, A
40	101.5	5.4	4545	2	US-08-804-227C-14	Sequence 14, Appli
41	101.5	5.4	4550	2	US-08-804-227C-18	Sequence 8, Appli
42	101.5	5.4	4550	2	US-08-804-198-2	Sequence 2, Appli
43	101	5.4	496	4	US-09-252-991A-31639	Sequence 31639, A
44	100.5	5.3	625	4	US-09-252-991A-23005	Sequence 23005, A
45	100.5	5.3	3491	2	US-07-642-734C-2	Sequence 2, Appli
46	100.5	5.3	3491	3	US-08-439-009A-2	Sequence 2, Appli
47	100	5.3	288	4	US-09-252-991A-23714	Sequence 23714, A
48	100	5.3	297	4	US-08-755-307A-3	Sequence 3, Appli
49	99.5	5.3	488	4	US-09-489-039A-13218	Sequence 13218, A
50	99.5	5.3	3170	3	US-09-036-987A-4	Sequence 4, Appli
51	99.5	5.3	3170	3	US-09-370-700-4	Sequence 4, Appli
52	99.5	5.3	3170	4	US-09-603-207-4	Sequence 4, Appli
53	99	5.3	443	4	US-09-252-991A-31911	Sequence 31911, A
54	98.5	5.2	379	4	US-09-252-991A-21117	Sequence 21117, A
55	98.5	5.2	986	4	US-08-311-731A-2	Sequence 2, Appli
56	98	5.2	381	4	US-09-489-039A-10807	Sequence 10807, A
57	98	5.2	513	4	US-09-252-991A-30773	Sequence 30773, A
58	98	5.2	1346	3	US-09-105-537-37	Sequence 37, Appli
59	98	5.2	1580	2	US-08-804-227C-11	Sequence 11, Appli
60	98	5.2	1580	2	US-08-804-198-5	Sequence 5, Appli
61	98	5.2	11877	3	US-09-105-537-6	Sequence 6, Appli
62	97.5	5.2	348	4	US-09-489-039A-10294	Sequence 10294, A
63	97.5	5.2	738	4	US-09-252-991A-30721	Sequence 30721, A
64	97	5.2	828	4	US-09-489-039A-12219	Sequence 12219, A
65	97	5.2	1151	4	US-09-252-991A-21328	Sequence 21328, A
66	96	5.1	550	4	US-09-252-991A-18287	Sequence 18287, A
67	96	5.1	679	4	US-09-252-991A-18681	Sequence 18681, A
68	96	5.1	3729	2	US-08-804-227C-4	Sequence 4, Appli
69	95.5	5.1	305	4	US-09-489-039A-9594	Sequence 9594, Ap
70	95.5	5.1	417	4	US-09-252-991A-28049	Sequence 28049, A
71	95.5	5.1	563	4	US-09-252-991A-31048	Sequence 31048, A
72	95.5	5.1	852	4	US-09-433-043B-118	Sequence 118, App
73	95	5.1	272	4	US-09-252-991A-26742	Sequence 26742, A
74	95	5.1	349	4	US-09-252-991A-31096	Sequence 31096, A
75	95	5.1	399	1	US-08-096-623A-20	Sequence 20, Appli
76	95	5.1	455	4	US-09-252-991A-17642	Sequence 17642, A
77	95	5.1	595	4	US-09-252-991A-30780	Sequence 30780, A
78	95	5.1	735	4	US-09-252-991A-30569	Sequence 30569, A
79	95	5.1	971	3	US-09-107-149-19	Sequence 19, Appli
80	95	5.1	1074	2	US-08-768-147B-2	Sequence 2, Appli
81	95	5.1	1074	3	US-08-968-752B-2	Sequence 2, Appli
82	95	5.1	1074	3	US-09-107-149-3	Sequence 3, Appli
83	95	5.1	1074	4	US-09-536-224-2	Sequence 2, Appli
84	95	5.1	4928	3	US-09-036-987A-5	Sequence 5, Appli
85	95	5.1	4928	3	US-09-370-700-5	Sequence 5, Appli
86	95	5.1	4928	3	US-09-603-207-5	Sequence 5, Appli
87	94.5	5.0	391	2	US-08-759-581B-13	Sequence 13, Appli
88	94.5	5.0	391	4	US-09-304-711-13	Sequence 13, Appli
89	94.5	5.0	391	4	US-09-173-281-13	Sequence 13, Appli
90	94.5	5.0	540	4	US-09-252-991A-19797	Sequence 19797, A
91	94.5	5.0	2035	1	US-08-046-585-5	Sequence 5, Appli
92	94.5	5.0	2035	1	PCT-US93-11721-5	Sequence 5, Appli
93	94.5	5.0	2035	1	US-08-393-703-3	Sequence 3, Appli
94	94.5	5.0	3739	3	US-09-320-678-2	Sequence 2, Appli
95	94.5	5.0	3739	3	US-09-105-537-33	Sequence 33, Appli
96	94.5	5.0	3739	4	US-09-141-908-3	Sequence 3, Appli
97	94.5	5.0	3739	4	US-09-657-440-2	Sequence 2, Appli
98	94	5.0	280	4	US-09-328-352-7042	Sequence 7042, Ap
99	94	5.0	711	1	US-08-235-638-7	Sequence 7, Appli
100	94	5.0	711	2	US-08-465-473B-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-08-984-618-14
; Sequence 14, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; APPLICANT: Tomasz, Alexander
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: NETHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: E. coli
; US-08-984-618-14

Query Match 95.9%; Score 1803; DB 3; Length 355;
Best Local Similarity 99.2%; Pred. No. 7.2e-179;
Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 2 MSGGKRLMVMAGTGGHVFPGGLAVAHLLMAQGWVRLGTADRMEDADLPKKGIEIDFI 61
Db 1 MSGGKRLMVMAGTGGHVFPGGLAVAHLLMAQGWVRLGTADRMEDADLPKKGIEIDFI 60
Qy 62 RISGLRGKIGKALIAAPLRFNARQARIMKAYKDPVILGMGGYVSGPGGLAAWSLGIP 121
Db 61 RISGLRGKIGKALIAAPLRFNARQARIMKAYKDPVILGMGGYVSGPGGLAAWSLGIP 120
Qy 122 VVLHEQNGIAGLTNKLARIATKVMQAPGAFNPAEYVGNPVRTDVLALPLPQORLAGRE 181
Db 121 VVLHEQNGIAGLTNKLARIATKVMQAPGAFNPAEYVGNPVRTDVLALPLPQORLAGRE 180
Qy 182 GPVRLVVGSGGARIINQNTMPQAAKLGDSVLIWHQSGKGSQSQSVQEAQVAGQPOHKV 241
Db 181 GPVRLVVGSGGARIINQNTMPQAAKLGDSVLIWHQSGKGSQSQSVQEAQVAGQPOHKV 240
Qy 242 TEFIDDDMAAAAYAWADVVCVRSGLTVSEIAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 301

Db 241 TEFIDDDMAAAAYAWADVVCVRSGLTVSEIAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 300
Qy 302 AKIIEQPQLSVDAVANTLAGWSRETLTMAERARAAASIPDATERVANEVSRVARA 356
Db 301 AKIIEQPQLSVDAVANTLAGWSRETLTMAERARAAASIPDATERVANEVSRVARA 355

RESULT 2

US-09-489-039A-12435
; Sequence 12435, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12435
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-12435

Query Match 86.6%; Score 1629; DB 4; Length 370;
Best Local Similarity 88.5%; Pred. No. 9.2e-161;
Matches 314; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
Qy 2 MSGGKRLMVMAGTGGHVFPGGLAVAHLLMAQGWVRLGTADRMEDADLPKKGIEIDFI 61
Db 15 MSGGKRLMVMAGTGGHVFPGGLAVAHLLMDQGWVRLGTADRMEDADLPKKGIEIDFI 74
Qy 62 RISGLRGKIGKALIAAPLRFNARQARIMKAYKDPVILGMGGYVSGPGGLAAWSLGIP 121
Db 75 LISGLRGKIGKALIAAPLRFNARQARIMKAYKDPVILGMGGYVSGPGGLAAWSLGIP 134
Qy 122 VVLHEQNGIAGLTNKLARIATKVMQAPGAFNPAEYVGNPVRTDVLALPLPQORLAGRE 181
Db 135 VVLHEQNGIAGLTNKLARIATKVMQAPGAFNPAEYVGNPVRTDVLALPLPQORLAGRE 194
Qy 182 GPVRLVVGSGGARIINQNTMPQAAKLGDSVLIWHQSGKGSQSQSVQEAQVAGQPOHKV 241
Db 195 GPVRLVVGSGGARIINQNTMPQAAKLGDSVLIWHQSGKGSQSQSVQEAQVAGQPOHKV 254
Qy 242 TEFIDDDMAAAAYAWADVVCVRSGLTVSEIAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 301
Db 255 TEFIDDDMAAAAYAWADVVCVRSGLTVSEIAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 314
Qy 302 AKIIEQPQLSVDAVANTLAGWSRETLTMAERARAAASIPDATERVANEVSRVARA 356
Db 315 AKIIEQPQLSVDAVANTLAGWSRETLTMAERARAAASIPDATERVANEVSRVARA 369

RESULT 3

US-09-543-681A-7067
; Sequence 7067, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7067
; LENGTH: 360
; TYPE: PRT

US-09-543-681A-7067

ORGANISM: Proteus mirabilis

Query Match 71.5%; Score 1345.5; DB 4; Length 360;
Best Local Similarity 72.1%; Pred. No. 2.5e-131;
Matches 259; Conservative 40; Mismatches 55; Indels 5; Gaps 1;

QY 2 MSQGGKRLMAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFI 61
DB 1 MSERKRLMAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIEI 60
QY 62 RISGLRGKIGKALIAAPLRIFNARWARARAIMKAYKDPVLMGMYGSGPGLAANSLGIP 121
DB 61 RISGLRGKIGKALIAAPLRIFNARWARARAIMKAYKDPVLMGMYGSGPGLAANSLGIP 120
QY 122 VVLEHQNGIAGLTKMLARIATKMQAEPGAFPAEAVVGNPRTDVLALPLPOORLAGRE 181
DB 121 VVLEHQNGIAGLTKMLARIATKMQAEPGAFPAEAVVGNPRTDVLALPLPOORLAGRE 180
QY 182 GPVRLVVGSGQGARILNQTMPQVAAKLGDSVLIWHQSGKSGQSQVEQAY----AEAGQ 236
DB 181 GAVRLVVGSGQGARILNQTMPQVAAKLGDSVLIWHQSGKSGQSQVEQAY----AEAGQ 240
QY 237 PQHKTBFIDDDMAAYAWADVVCRCGALTVEIAAGLPALFPVPPQHKDRQOYWNALPL 296
DB 241 SEYKVTFFIDDIAGYQWADVVCRCGALTVEIAAGLPALFPVPPQHKDRQOYWNALPL 300
QY 297 EKAGAAKIIQPOLSVDAVANTLAGRSRETLTMAERARAASIPDATERVANEVSVAR 355
DB 301 ENAGAARIIEQNDLTPEAIATDLENWDRHQLMLAKQAQVAITDATERVANVIEIVAK 359

RESULT 4

US-08-984-618-15

Sequence 15, Application US/08984618

Patent No. 6251647

GENERAL INFORMATION:

APPLICANT: de Lencastre, Herminia

APPLICANT: Tomasz, Alexander

TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF

TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/984,618

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCES/DOCKET NUMBER: 600-1-213 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 351 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Haemophilus influenzae

US-08-984-618-15

Query Match 56.6%; Score 1064.5; DB 3; Length 351;
Best Local Similarity 61.5%; Pred. No. 3.8e-102;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;

QY 5 QCKRLMWAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFIRIS 64
DB 2 KKKLLMWAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFIRIS 61
QY 65 GLRGKIGKALIAAPLRIFNARWARARAIMKAYKDPVLMGMYGSGPGLAANSLGIPVL 124
DB 62 GLRGKIGKALIAAPLRIFNARWARARAIMKAYKDPVLMGMYGSGPGLAANSLGIPVL 121
QY 125 HEONGIAGLTKMLARIATKMQAEPGAFPAEAVVGNPRTDVLALPLPOORLAGREGPV 184
DB 122 HEONGIAGLTKMLARIATKMQAEPGAFPAEAVVGNPRTDVLALPLPOORLAGREGPV 181
QY 185 RVLVVGSGQGARILNQTMPQVAAKLGDSVLIWHQSGKSGQSQVEQAYAEAGQPOHKTVEF 244
DB 182 RVLVVGSGQGARILNQTMPQVAAKLGDSVLIWHQSGKSGQSQVEQAYAEAGQPOHKTVEF 240
QY 245 IDDMAAAYAWADVVCRCGALTVEIAAGLPALFPVPPQHKDRQOYWNALPLEKAGAAKI 304
DB 241 IDDMAAAYAWADVVCRCGALTVEIAAGLPALFPVPPQHKDRQOYWNALPLEKAGAAKI 300
QY 305 IEQPOLSVDAVANTLAGRSRETLTMAERARAASIPDATERVANEVS 352
DB 301 IEQADLTPEILVNLKNTRENLLQALKAKTMSWPNAAQORVAEVIKQ 348

RESULT 5

US-09-252-991A-24124

Sequence 24124, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 24124

LENGTH: 390

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-24124

Query Match 38.0%; Score 715.5; DB 4; Length 390;
Best Local Similarity 44.2%; Pred. No. 8.2e-66;
Matches 157; Conservative 64; Mismatches 123; Indels 11; Gaps 6;

QY 9 LMVWAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFIRISGLRG 68
DB 38 VLMWAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFIRISGLRG 97
QY 69 KGKALIAAPLRIFNARWARARAIMKAYKDPVLMGMYGSGPGLAANSLGIPVVLHEQN 128
DB 98 KGKSLVAPLELLKSLFQALRVIRQLRVCVLGLGYYVIGPGLAARLNGVPLVHEQN 157
QY 129 GIAGLTKMLARIATKMQAEPGAFPAEAVVGNPRTDVLALPLPOORLAGREGPV 185
DB 158 AVAGTANRSLAPITARVCEAFDPTFPASDKRLTGTGNVPGELFLDAHARAPLTGR--VN 215
QY 186 RVLVVGSGQGARILNQTMPQVAAKLGDSVLIWHQSGKSGQSQVEQAYAEAGQPOHKTVE 243

Db 9 KILVTGGTGGHHYIPALSFVHVKEAPATBFLYVGTENGLESQIVPKKIPFKTIKIQG 68
Qy 66 ----LRGKIGIKALIAAPLRFNARQARAIKAYKPDVVLGMGGVSPGGLAAMSLGIP 121
Db 69 FKRSLSQPNFKTIYL----FLTSINKAKKIIREFQPDVVIKGGVSGAVVAAHQLKIP 124
Qy 122 VVLHEQNGIAGLTNKLARIATKVMQABPGA---PPNAEVV--GNPVRTDVLALPLPQOR 176
Db 125 TIIHQNSIPGNTNKLRSYVDKIAICFPDVASFPFKETILTGNRGOEVVT--VEKSA 182
Qy 177 LAGREG--PVR--VLVVGSGQARILNQTMPQVAAKLG--SVIIHQSGKGSQSQVQEA- 230
Db 183 ILSEFGLDPAKTVVLFGSRGALKINQAFQAPLFEEREYQVLYASGERYYQELQESL 242
Qy 231 -YAEAGQPQHKVTEFIDDMAAYANADVVVCRSGALTVSEIAAAGLPAFVFPQH-KDRQ 288
Db 243 KSEKKLTNISVQPYIDKMWVNMANTDLMVGRAGTSAEFTALGLPALIPSPYVINDH 302
Qy 289 QYWNALPLEKAGAKKIIQEPQLS----VDVANTLAGWSRETLTMAERARAAISIPDATE 344
Db 303 QTKNAQSLVKVGAVEMIPDAELTGARLVAIDILL--NNEKQQAATASKGERIPDASD 360
Qy 345 RV 346
Db 361 RL 362

RESULT 11
US-08-984-618-16
; Sequence 16, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; ADDRESSEE: de Lencastre, Herminia
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecalis
; US-08-984-618-16

Query Match 21.6%; Score 406.5; DB 3; Length 363;
Best Local Similarity 30.9%; Pred. No. 9.4e-34;
Matches 112; Conservative 75; Mismatches 144; Indels 31; Gaps 13;
Qy 8 RLMVMAGGTGGHVFPGLAVAHLM--AQGWQVRWLTADRMADLVPKHGIEIDFIRISG 65
Db 2 KILVTGGTGGHHYIPALSFVHVKEAPATBFLYVGTENGLESQIVPKKIPFKTIKIQG 61
Qy 66 ----LRGKIGIKALIAAPLRFNARQARAIKAYKPDVVLGMGGVSPGGLAAMSLGIP 121
Db 62 FKRSLSQPNFKTIYL----FLTSINKAKKIIREFQPDVVIKGGVSGAVVAAHQLKIP 117
Qy 122 VVLHEQNGIAGLTNKLARIATKVMQABPGA---PPNAEVV--GNPVRTDVLALPLPQOR 176
Db 118 TIIHQNSIPGNTNKLRSYVDKIAICFPDVASFPFKETILTGNRGOEVVT--VEKSA 175
Qy 177 LAGREG--PVR--VLVVGSGQARILNQTMPQVAAKLG--SVIIHQSGKGSQSQVQEA- 230
Db 176 ILSEFGLDPAKTVVLFGSRGALKINQAFQAPLFEEREYQVLYASGERYYQELQESL 235
Qy 231 -YAEAGQPQHKVTEFIDDMAAYANADVVVCRSGALTVSEIAAAGLPAFVFPQH-KDRQ 288
Db 236 KSEKKLTNISVQPYIDKMWVNMANTDLMVGRAGTSAEFTALGLPALIPSPYVINDH 295
Qy 289 QYWNALPLEKAGAKKIIQEPQLS----VDVANTLAGWSRETLTMAERARAAISIPDATE 344
Db 296 QTKNAQSLVKVGAVEMIPDAELTGARLVAIDILL--NNEKQQAATASKGERIPDASD 353
Qy 345 RV 346
Db 354 RL 355
RESULT 12
US-09-198-452A-977
; Sequence 977, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 977
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; US-09-198-452A-977
Query Match 15.1%; Score 284.5; DB 4; Length 357;
Best Local Similarity 29.5%; Pred. No. 4.4e-21;
Matches 104; Conservative 59; Mismatches 142; Indels 47; Gaps 14;
Qy 7 KRLMVMAGGTGGHVFPGLAVAHLM--AQGWQVRWLTADRMADLVPKHGIEIDFIRISGL 66
Db 6 RKVALVGGSGGHHYIPALSFVHVKEAPATBFLYVGTENGLESQIVPKKIPFKTIKIQG 62
Qy 67 RG--KGKIKALIAAPLRFNARQARAIKAYKPDVVLGMGGVSPGGLAAMSLGIPVVL 124
Db 63 PTVLNPIK-IMSRTLSCSGYLKARKELKIFDDVLIGFSGYHSLPVLLAGLSHKIPFL 121
Qy 125 HEQNGIAGLTNKLARIATKVMQABPGA---PPNAEVV--GNPVRTDVLALPLPQOR 174
Db 122 HEQNLVFGKYNQLFSRYARGVGNFSPVTKHFRCP-----ABEVFLPKES--FSLGSPM 173
Qy 175 ORLAGREGPVRVLVVGSGQARILNQTMPQVAAKLG--SVIIHQSG--KGSQSQVQEA 230
Db 174 MKRCTNHTPT-ICVVGSGQAQILNCTVPOALVKLVNKNYPNMYHHVIGPKSUMVQHV 232
Qy 231 YAEAGQPQHKVTEFIDDMAAYANADVVVCRSGALTVSEIAAAGLPAFVFPQH-KDRQ 290

Db 233 Y-NRGEVLCVCKPFEQQLDLVLAADLVISRAGATILEILWAKVPGLIPPGAYGHQE 291
Qy 291 WNA-----LPEKAGAAK-IIEOPQLSVDA-----VANTLAGMSRE 325
Db 292 VNAKFFVDVLEGGTMIKELEKTEKLIVKVTFALDSHNEKQNSLAAYSQQ 343
RESULT 13
US-08-751-474-2
; Sequence 2, Application US/08751474
; Patent No. 5821335
; GENERAL INFORMATION:
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Skatrud, Paul L.
; TITLE OF INVENTION: Biosynthetic Gene murG From
; TITLE OF INVENTION: Streptococcus Pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,474
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9902
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-474-2
Query Match 14.5%; Score 273.5; DB 2; Length 352;
Best Local Similarity 27.7%; Pred. No. 5.9e-20;
Matches 93; Conservative 52; Mismatches 150; Indels 41; Gaps 8;
Qy 7 KRLVMVAGGTGGHVPFGLAVAHILMAQGWVRLGTADRMEDLVPKHGIEIDFIRISGL 66
Db 2 KXIVFTGGTGVHTVTLNLLMKPFIEDGWEVHYIGDKRGIEHQEILASGLDVTFHSIA-- 59
Qy 67 RGKGIKALIAAPLIRFNAR-----QARAIMKAYKPDVVLGMGVVSGPGGL 113
Db 60 TGG-----LRRYFSNQNMFLDKVKGWGVQSLFIMLRPOTLFSKGGFVSPFVI 110
Qy 114 AAWSLGIPVVLHEQNGIAGLTNKLARATKVMQA--EPGAFNAEVVGNPVRTDLALP 171
Db 111 AARVSGVFPVTHESDLSGLANKIAYKATKMYSTFEQASLSKVHGVATKVSQDNKP 170
Qy 172 LPQ-----QLRAGRGVPRVLVVGSGQARILNQTMPQVAAKLGDSVLIHQSGKGSQS 226
Db 171 EDELVDVDTQTHFNHKLPT-VLFVGGSGARVFNQLVTDKHELTERYNIINLTGDSLIN- 228
Qy 227 EQVAAEAGQPKHKTETIDDMAAYAWADVVCRSGLTVEIAAAGLPALFVFPQHKD 286
Db 229 -----ELSQNLFRVDYTDLYQPLMELADVVVTRGGATIFELLATAKLHVIVELGREA 282
Qy 287 RQ--QYWNALPLEKAGAAKIIIEOPQLSVDAVANTLA 320

Db 283 SRGQIENAYFVKGYAEDLQESDULDSLEKLS 318
RESULT 14
US-09-134-001C-3198
; Sequence 3198, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3198
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3198
Query Match 13.9%; Score 261; DB 4; Length 366;
Best Local Similarity 28.1%; Pred. No. 1.3e-18;
Matches 92; Conservative 55; Mismatches 140; Indels 40; Gaps 11;
Qy 8 RLMVMAGTGGHVPFGLAVAHILMAQGWVRLGTADRMEDLVPKHGIEIDFIRISGLR 67
Db 12 KIAYTGGTGVHVSVNLSTIEKGEAFYIGSKGIEREMIESQLPDIOYVPFIS-- 69
Qy 68 GGIKAL-----IAAPLIRFNARQARAIMKAYKPDVVLGMGVVSGPGGLAAWSLIGIPW 123
Db 70 GKRLYSFENAKDVFKVLKGLDARKILKKQKDLLEFSKGGFVSPVVIARSLKIPTI 129
Qy 124 LHEQNGIAGLTNKLARATKVMQAEPGAF-----PNAEVVGNPVRTDLALPLPQOLA 178
Db 130 IHESDLTPGLANKISLTKFAKIYTTFTDTLVLPKDKADFVGATVREDL-----KQG 181
Qy 179 GREGPVR-----VLVVGSGGARILNQTMPQ--VAAKLGDSVLIHQSGKGSQSV 227
Db 182 NKEGYQLTDFDNKKVLLVMGSGSLGKLNINIRQNTAEALLHDYQII-HLTGKG--LV 237
Qy 228 EQVAAEAGQPKHKTETI--DMAAYAWADVVCRSGLTVEIAAAGLPALFVP--FQH 284
Db 238 DSIINKKGVQP---EFVKDDLTLDLAITDTVVSRAAGNAIVEFLSLRIPMLLIPLGLDQ 294
Qy 285 KDRQQYWNALPLEKAGAAKIIIEOPQLS 311
Db 295 SRGQDQIDNAKNFESKGYGRHIPEDQLT 321
RESULT 15
US-08-984-618-17
; Sequence 17, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; APPLICANT: Tomas, Alexander
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,618
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-213 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Cucumis sativus
US-08-984-618-17

Query Match 7.0%; Score 131; DB 3; Length 525;
Best Local Similarity 20.4%; Pred. No. 7e-05;
Matches 66; Conservative 55; Mismatches 121; Indels 82; Gaps 14;
QY 82 FNAWEQARAIKAYKPDVVLGMGVGSGGLAAMSLGIPVVLHQNQI---AGLTNKLW 138
DB 224 FIAREVAKGLMK-YRPDIISVH-----FLMQHPVPIRLASKGLLNKIV 266
QY 139 -ARIATKVNQAPGAF-----PNAEVVG-----NPVTDVLAFLPLP----- 173
DB 267 FTTVVTDLSTCHPTWTFHKLIVTRCYCPSTEVAKATAGLQPSKLKVLGFLVPSPVKPIR 326
QY 174 ----CORLAGREGVRVLVWGSSQCARILNQTMPQVAAKLGDS-----VILWHQ 218
DB 327 PKIELRKELGMENLPAVLLMGSGEGMPTEATKALSALYDENHGBPIGQVLVICGN 386
QY 219 SG-KGQQSVQAYAEAGQPHKVTETFDIDMAAAYANADVVCVRSALTVEIAAAGLPA 277
DB 387 KKLAGRLRSID-----WKVPVQKGFVTMBECCMGACDCIITKAGPGGTIAEAMIRGLPI 440
QY 278 L----FVPPQHKDQYWNALPLEKAGAAKIIQPOLSVDAVANTLAGW---SRETLTWA 331
DB 441 ILNDYIAGQAGNVPY-----VVENGCGFKSKSPK-----EIANIVAKWFGPKADELLIMS 491
QY 332 ERARAASIPDATERVANEVSRVAR 355
DB 492 QNALRLARPDAVFKIVHDLHELK 515

RESULT 16
US-08-984-618-12
Sequence 12, Application US/08984618
Patent No. 6251647
GENERAL INFORMATION:
APPLICANT: de Lencastre, Herminia
APPLICANT: Tomasz, Alexander
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA

ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,618
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-213 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus subtilis
US-08-984-618-12
Query Match 6.8%; Score 127; DB 3; Length 382;
Best Local Similarity 18.9%; Pred. No. 0.00011;
Matches 59; Conservative 65; Mismatches 118; Indels 70; Gaps 12;
QY 80 RIFNAM-----ROARAIKAYKPDVVLGMGVGSGGLAAMSLGIP-----V 122
DB 83 RKENIYFMGNKGLGELVDEHQDIIINTFPMIIVPEYRRTGRVPTFNVTDFCLHKI 142
QY 123 VLHEQNGIAGLTNKLARIATKVAQEPGAF--NAEVGPNVPTDVLALPLQORLAGRE 181
DB 143 WVHE-----NVDKYIATDYVKEKLEIGHFNSVKAITGIPR-----PQFESMPV 189
QY 182 GPVR-----VLVVGSSQGA-----RLNQTMPQVAAKLGDSVILHQSCK 221
DB 190 GPIYKYNLSPNKKVLLINAGAHGVLNKKELCNLVKDDQVQVYVCGKNTAL----- 243
QY 222 GSQSVQAYAEAGQPHKVTETFDIDMAAAYANADVVCVRSALTVEIAAAGLPA 279
DB 244 --XESLSALEAENG-KLKVLYGVVERIDELFRITDCMITKPGGITTLEATAIGPVILYK 300
QY 280 -VPFQHKDQYWNALPLEKAGAAKIIQPOLSVDAVANTLAGWRETLTMAERARAAS 338
DB 301 FVPGQEK-----NANFFEDRGAAIVVNRHEETLESVTSLLA--DEDTLHRMKNKINDLH 353
QY 339 IPDATERVANEV 350
DB 354 LANSSEVILEDI 365

RESULT 17
US-09-252-991A-26203
Sequence 26203, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 28203
;; LENGTH: 572
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26203

Query Match 6.5%; Score 122.5; DB 4; Length 572;
Best Local Similarity 23.8%; Pred. No. 0.00061;
Matches 93; Conservative 47; Mismatches 154; Indels 97; Gaps 17;
QY 18 GHVPPGLAVAHLMQAQ---GQVRLGTADMEADLVPHGIBIDFIRIS-GLRGKGIKA 73
Db 157 GYVLADLEQGRFVADLLGGQVVEGAADWLDAAARLPDPEAALAHVTPSARAPRAE 216
QY 74 LIAAPLRFIFNWRQARAIMKAYKEDVV-----LGMGGYVSGGGL-----AANSLGI 120
Db 217 LLIHPRCVLALEPADAAADAVRQALVDAGLAAALGGLLADARSADPRLQAAADALGV 276
QY 121 PVVLHEQNGIAGLTNKLARIATKVMQAEFG-----APFNAEVVGNPVRTDVLALPLPQQ 175
Db 277 PLRF-----LPOADPAARLRQALPRAEQVGD--RLAVAAAPLAMD 314
QY 176 RLACREGBRVVLVVGSGGARILNQTWPQVAAKLGDSVIIHQSGKGSQSQVEQAYA-EA 234
Db 315 RIGORRG--RLTVIGLGGCAELMVPAARRALDEAEDIL-----GYQTYVDMAGPFRA 365
QY 235 GQPKHVTFFDDMAAVAW-----ADVVCVRSGALTVSEIAAGLPAFPVPPQHKDRQ 288
Db 366 GOVLHSDNREELQARHAFELATGRVAVLVVSGDPCGVFAMAAVLEAL-----HGAGE 420
QY 289 QYWNALPLE-----KAGAKIIEQP---QLSVDVANTLAGSRETLLTMAE 332
Db 421 PANDAVELRIFPGVSAALATAARAGA-----PLGHPFCVLSLSDNLKPWE-----VIER 469
QY 333 RARASIPDATERVANEVSR-----VARALE 358
Db 470 RDLAGAADLAWAFNPISRARPWOLGRALE 500

RESULT 18

US-08-311-731A-7
; Sequence 7, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/311,731A
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 463 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM TUBERCULOSIS
US-08-311-731A-7

Query Match 6.2%; Score 117; DB 4; Length 463;
Best Local Similarity 27.9%; Pred. No. 0.0016;
Matches 46; Conservative 28; Mismatches 51; Indels 40; Gaps 7;
QY 203 POVAAKLGDSVIIHQSGKGSQSQVEQAYAE-----AGQPQHK-----VTEFID 246
Db 267 FIIVATLSS-----GGKNLLQVVLNALADLPVTVIATAGRNHLKNVSPANAFVADYLP 320
QY 247 DMAAAYAVADVVCVRSGALTVSEIAAGLPAFPVPPQHKDRQYWNALPLEKAGAKIIE 306
Db 321 GEAAAARLA-VVLCNGGSPTTQCALAAGVPVIGLP---SNMDQHLNMEALERAGAGVILLR 376
QY 307 QPQLSVDVANTLAGSRETLLTMAERARAA-----SIPDA 342
Db 377 TERLNTGVAARV-----KQVLSGAEFRQARRLPKPSDQILPDS 416

RESULT 19

US-09-036-987A-8
; Sequence 8, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R.
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 390 amino acids
; TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-036-987A-8

Query Match 6.1%; Score 115.5; DB 3; Length 390;
Best Local Similarity 20.6%; Pred No. 0.0018;
Matches 85; Conservative 46; Mismatches 126; Indels 155; Gaps 20;
QY 14 GGTGGHV---FPGLAVAHLMQAQGVWLTADRM---EADLVPKHGIEIDFIRISGLR 67
DB 60 GDTGGTTQLRFPNPAFGQRTAGRQL-WEQTASNVAAQSSLDOLPEY-----105
QY 68 GKGKALIAAPLRIFNARQARIMKAYKPDVVLGMGVSGPGGLAAWSLGIPIWLH--125
DB 106 -----LRLAEAWRPSVLLV-----DVCALIGRVL-----LDLPVVLHRW 142
QY 126 -----EQNGIAGLTNKLARIATKVMQAEPPGAPFNAEVVGNPVR 164
DB 143 GVDPTAGPFSDBRAHELDPVCRHHGLTGL-----PTPELILDP 182
QY 165 TDVLALPLPQORLAGREGPVRVLLVVGGS-----QGARILNQTMP--203
DB 183 PSLOASDAQ-----GAPVQVYPYNGSGAFPAGWAARTSARRVCICMGRWLNATGPAP 236
QY 204 ---QVAAKL---GDSVVIHQSGKSGSQSVEQAYAEAGOPQH--VTEFIDDMAAAYA--253
DB 237 LIRAVAAATELPGVEAVI-----AVPPEHRALLTDLDPNARIAESVP 278
QY 254 -----WADVVCRSALTVSIEIAAGLPALFPVPHQKDRQOYWNALPLEKAGAKIIEQ 307
DB 279 LNLFLRTCELVICAGSGTAFTATRLGIPQLVLP-QYFDQFDYARNLAAAGAGICLPDEQ 337
QY 308 POLS-----VDVAVANTL---AGWSRETLTMAERARAASIPDATERVANEVSRV 353
DB 338 AQSDHEQFTDSIATVLGDTGFA-SAAIKLSDEITAMPHPAALVRLTLENTAAI 388

RESULT 20

US-09-370-700-8
Sequence 8, Application US/09370700
Patent No. 6274350

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H
APPLICANT: Broughton, Mary C
APPLICANT: Crawford, Kathryn P
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patti J
APPLICANT: Turner, Jan E
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentin Ver..2.0
SEQ ID NO 8
LENGTH: 390
TYPE: PRT
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-8

Query Match 6.1%; Score 115.5; DB 3; Length 390;
Best Local Similarity 20.6%; Pred. No. 0.0018;
Matches 85; Conservative 46; Mismatches 126; Indels 155; Gaps 20;

QY 14 GGTGGHV---FPGLAVAHLMQAQGVWLTADRM---EADLVPKHGIEIDFIRISGLR 67
DB 60 GDTGGTTQLRFPNPAFGQRTAGRQL-WEQTASNVAAQSSLDOLPEY-----105
QY 68 GKGKALIAAPLRIFNARQARIMKAYKPDVVLGMGVSGPGGLAAWSLGIPIWLH--125

DB 106 -----LRLAEAWRPSVLLV-----DVCALIGRVL-----LDLPVVLHRW 142
QY 126 -----EQNGIAGLTNKLARIATKVMQAEPPGAPFNAEVVGNPVR 164
DB 143 GVDPTAGPFSDBRAHELDPVCRHHGLTGL-----PTPELILDP 182
QY 165 TDVLALPLPQORLAGREGPVRVLLVVGGS-----QGARILNQTMP--203
DB 183 PSLOASDAQ-----GAPVQVYPYNGSGAFPAGWAARTSARRVCICMGRWLNATGPAP 236
QY 204 ---QVAAKL---GDSVVIHQSGKSGSQSVEQAYAEAGOPQH--VTEFIDDMAAAYA--253
DB 237 LIRAVAAATELPGVEAVI-----AVPPEHRALLTDLDPNARIAESVP 278
QY 254 -----WADVVCRSALTVSIEIAAGLPALFPVPHQKDRQOYWNALPLEKAGAKIIEQ 307
DB 279 LNLFLRTCELVICAGSGTAFTATRLGIPQLVLP-QYFDQFDYARNLAAAGAGICLPDEQ 337
QY 308 POLS-----VDVAVANTL---AGWSRETLTMAERARAASIPDATERVANEVSRV 353
DB 338 AQSDHEQFTDSIATVLGDTGFA-SAAIKLSDEITAMPHPAALVRLTLENTAAI 388

Search completed: June 7, 2004, 07:15:50
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 07:14:21 ; Search time 47 Seconds
(without alignments)
2178.878 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 MYSQGRMLVMAGTGGHV.....RVANEVSRVALDEHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*

17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1881	100.0	364	10	US-09-829-275-1
2	1803	95.9	355	9	US-09-741-669-454
3	1803	95.9	355	9	US-09-815-242-10031
4	1803	95.9	355	12	US-10-282-122A-56415, A
5	1687	89.7	355	12	US-10-282-122A-75189
6	1677	89.2	355	9	US-09-815-242-14078
7	1677	89.2	355	12	US-10-282-122A-76021
8	1549.5	85.3	348	12	US-10-282-122A-59581
9	1549.5	82.4	348	12	US-10-282-122A-73040
10	1471	78.2	356	12	US-10-282-122A-77926
11	1345.5	71.5	360	12	US-10-282-122A-68811
12	1323	70.3	281	12	US-10-282-122A-55838
13	1154	61.4	354	12	US-10-282-122A-77456
14	1123	59.7	354	12	US-10-282-122A-66899
15	1064.5	56.6	351	9	US-09-815-242-11189

16	1064.5	56.6	351	12	US-10-282-122A-58422
17	1061.5	56.4	351	10	US-09-829-275-2
18	871	46.3	367	12	US-10-282-122A-50225
19	849	45.1	372	12	US-10-282-122A-49098
20	791	42.1	357	12	US-10-282-122A-51374
21	772.5	41.1	355	12	US-10-282-122A-66061
22	759.5	40.4	355	12	US-10-282-122A-65125
23	719.5	38.3	357	9	US-09-815-242-12009
24	719.5	38.3	357	12	US-10-282-122A-66651
25	717	38.1	356	12	US-10-282-122A-69709
26	696.5	37.0	367	12	US-10-282-122A-63072
27	684	36.4	375	12	US-10-282-122A-68222
28	672.5	35.8	365	12	US-10-282-122A-44666
29	535	28.4	408	12	US-10-282-122A-62025
30	534	28.4	250	12	US-10-282-122A-47797
31	510.5	27.1	407	12	US-10-282-122A-63788
32	506	26.9	410	9	US-09-712-363-223
33	506	26.9	410	12	US-10-282-122A-62331
34	506	26.9	410	12	US-10-282-122A-64551
35	500	26.6	410	10	US-09-829-275-8
36	488.5	26.0	359	12	US-10-282-122A-54021
37	487	25.9	372	9	US-09-738-626-5863
38	479	25.5	380	12	US-10-282-122A-48824
39	470.5	25.0	409	12	US-10-282-122A-53422
40	461.5	24.5	363	14	US-10-156-761-13649
41	460	24.5	383	12	US-10-282-122A-45908
42	423.5	22.5	366	12	US-10-282-122A-46233
43	422	22.4	363	12	US-10-282-122A-60776
44	420	22.3	362	12	US-10-282-122A-57874
45	419	22.3	360	10	US-09-829-275-4
46	418	22.2	363	10	US-09-829-275-7
47	411.5	21.9	363	12	US-10-282-122A-42564
48	406.5	21.6	362	10	US-09-829-275-3
49	402.5	21.4	368	9	US-09-815-242-4866
50	401.5	21.3	363	9	US-09-815-242-4866
51	335.5	17.8	354	12	US-10-282-122A-51853
52	328.5	17.5	359	12	US-10-282-122A-76528
53	315.5	16.8	384	12	US-10-282-122A-76528
54	313.5	16.7	385	10	US-09-829-275-6
55	303	16.1	353	12	US-10-282-122A-55084
56	284.5	15.1	357	12	US-10-282-122A-54956
57	284.5	15.1	357	15	US-10-312-273-107
58	284.5	15.1	357	15	US-10-289-762-977
59	277	14.7	353	9	US-09-815-242-11564
60	277	14.7	353	12	US-10-335-977-7604
61	277	14.7	363	12	US-10-335-977-7605
62	275.5	14.6	352	9	US-09-815-242-13634
63	275	14.6	353	12	US-10-282-122A-58892
64	273	14.5	356	9	US-09-815-242-13614
65	273	14.5	356	9	US-09-815-242-12853
66	272.5	14.5	356	12	US-10-282-122A-70334
67	272.5	14.5	352	9	US-09-815-242-13399
68	272.5	14.5	352	10	US-09-829-275-5
69	272.5	14.5	352	12	US-10-282-122A-73866
70	263	14.0	342	12	US-10-282-122A-54490
71	261	13.9	357	12	US-10-282-122A-70571
72	259.5	13.8	363	12	US-10-282-122A-47259
73	250	13.3	360	12	US-10-282-122A-74653
74	213.5	11.4	361	12	US-10-282-122A-72041
75	160.5	8.5	186	12	US-10-282-122A-71400
76	149	7.9	397	14	US-10-156-761-14886
77	129	6.9	453	12	US-10-424-599-231702
78	129	6.9	454	12	US-10-425-114-55430
79	122	6.5	393	12	US-10-425-114-46369
80	122	6.5	511	12	US-10-425-114-65133
81	120	6.4	453	15	US-10-369-493-562
82	118	6.3	516	12	US-10-424-599-178191
83	115.5	6.1	397	16	US-10-329-148A-8
84	113.5	6.0	460	12	US-10-425-114-39058
85	113	6.0	412	12	US-10-425-114-47475
86	113	6.0	652	14	US-10-156-761-13141
87	111	5.9	399	12	US-10-107-431-223
88	111	5.9	399	12	US-09-758-759-9

Sequence 58422, A
Sequence 2, Appli
Sequence 50225, A
Sequence 49098, A
Sequence 51374, A
Sequence 60661, A
Sequence 65125, A
Sequence 12009, A
Sequence 66651, A
Sequence 69709, A
Sequence 63072, A
Sequence 68222, A
Sequence 44666, A
Sequence 60205, A
Sequence 47797, A
Sequence 63788, A
Sequence 223, App
Sequence 62331, A
Sequence 64651, A
Sequence 8, Appli
Sequence 54021, A
Sequence 5863, Ap
Sequence 48824, A
Sequence 53422, A
Sequence 13649, A
Sequence 45908, A
Sequence 46233, A
Sequence 60776, A
Sequence 57874, A
Sequence 4, Appli
Sequence 7, Appli
Sequence 42564, A
Sequence 10637, A
Sequence 3, Appli
Sequence 4866, Ap
Sequence 52804, A
Sequence 51853, A
Sequence 76528, A
Sequence 6, Appli
Sequence 55084, A
Sequence 54956, A
Sequence 107, App
Sequence 977, App
Sequence 11564, A
Sequence 7604, Ap
Sequence 7605, Ap
Sequence 13634, A
Sequence 58892, A
Sequence 12614, A
Sequence 12853, A
Sequence 70334, A
Sequence 13399, A
Sequence 5, Appli
Sequence 73866, A
Sequence 54490, A
Sequence 70571, A
Sequence 47259, A
Sequence 74653, A
Sequence 72041, A
Sequence 71400, A
Sequence 14886, A
Sequence 231702, A
Sequence 55430, A
Sequence 46369, A
Sequence 65133, A
Sequence 562, App
Sequence 178191, A
Sequence 8, Appli
Sequence 39058, A
Sequence 47475, A
Sequence 13141, A
Sequence 223, App
Sequence 9, Appli

89 110.5 5.9 727 16 US-10-408-765A-636 Sequence 636, App
90 110.5 5.9 10421 12 US-10-282-122A-61631 Sequence 61631, A
91 109.5 5.8 4471 14 US-10-205-032-10 Sequence 10, Appl
92 108 727 16 US-10-408-765A-1646 Sequence 1646, Ap
93 107.5 5.7 240 12 US-10-424-599-150648 Sequence 150648, A
94 107.5 5.7 412 14 US-10-156-761-8483 Sequence 43991, A
95 107 5.7 235 12 US-10-425-114-43991 Sequence 2, Appl
96 106.5 5.7 168 14 US-10-360-053-2 Sequence 4, Appl
97 106.5 5.7 168 14 US-10-360-053-4 Sequence 50, Appl
98 106.5 5.7 590 16 US-10-220-480-50 Sequence 56, Appl
99 106.5 5.7 612 16 US-10-220-480-56 Sequence 10834, A
100 106.5 5.7 1078 14 US-10-156-761-10834

ALIGNMENTS

RESULT 1
US-09-829-275-1
; Sequence 1, Application US/09829275
; Publication No. US20030077803A1
; GENERAL INFORMATION:
; APPLICANT: WALKER, SUZANNE
; TITLE OF INVENTION: CRYSTALS OF THE ESCHERICHIA COLI MEMBRANE-ASSOCIATED
; TITLE OF INVENTION: GLYCOSYLTRANSFERASE (MURG) PROTEIN, ATOMIC COORDINATES
; TITLE OF INVENTION: AND THREE DIMENSIONAL STRUCTURES THEREOF, ATOMIC
; TITLE OF INVENTION: COORDINATES AND THREE DIMENSIONAL STRUCTURES OF
; TITLE OF INVENTION: BINDING DOMAINS THEREOF, IMAGES THEREOF, AND METHODS
; TITLE OF INVENTION: OF CRYSTALLIZING MURG PROTEINS, MODELS OF UDP-
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES, MURG PROTEINS AND BINDING SITES,
; TITLE OF INVENTION: METHODS OF MAKING MODELS, METHODS OF USING MODELS OF
; TITLE OF INVENTION: MURG, COMPOUNDS THAT BIND, INHIBIT OR STIMULATE MURG
; TITLE OF INVENTION: PROTEINS, AND THERAPEUTIC COMPOSITIONS THEREOF
; FILE REFERENCE: 4555-105
; CURRENT APPLICATION NUMBER: US/09/829,275
; CURRENT FILING DATE: 2001-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-829-275-1
Query Match 100.0%; Score 1881; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 1.7e-173; Indels 0; Gaps 0;
Matches 364; Conservative 0; Mismatches 0
Qy 1 MMSGQGKRLVMVAGGTGGHVFPGLAVAHLMAGQVQWRLGTADRMADLVPKHGIEIDF 60
Db 1 MMSGQGKRLVMVAGGTGGHVFPGLAVAHLMAGQVQWRLGTADRMADLVPKHGIEIDF 60
Qy 61 IRISGLRGKIKALIAAPLRIFNARQARIMKAYKPDVVLGNGGYVSGPGGLAANSLGI 120
Db 61 IRISGLRGKIKALIAAPLRIFNARQARIMKAYKPDVVLGNGGYVSGPGGLAANSLGI 120
Qy 121 PVVLEHQNGIAGLTNKLARIATKVMQAEPCAFNVAEVVGNPVRTDVLALPLPQORLAGR 180
Db 121 PVVLEHQNGIAGLTNKLARIATKVMQAEPCAFNVAEVVGNPVRTDVLALPLPQORLAGR 180
Qy 181 EGPVRLVVGSGQCARILNQTMPQVAAKLGDSVTIWHQSGKGSQQSVQVQYAEAGQPHK 240
Db 181 EGPVRLVVGSGQCARILNQTMPQVAAKLGDSVTIWHQSGKGSQQSVQVQYAEAGQPHK 240
Qy 241 VTEFIDMAAYAWADVVCVCSGALTIVSEIAAAGLPALFVFPQHKDQOQYVNALPLEKAG 300
Db 241 VTEFIDMAAYAWADVVCVCSGALTIVSEIAAAGLPALFVFPQHKDQOQYVNALPLEKAG 300
Qy 301 AAKTIEQPQLSVDAVANTLAGSRETLTMAERARAASIPDATERVANEVSRVARALEHH 360
Db 301 AAKTIEQPQLSVDAVANTLAGSRETLTMAERARAASIPDATERVANEVSRVARALEHH 360
Qy 361 HHHH 364

Db 361 HHHH 364
RESULT 2
US-09-741-669-454
; Sequence 454, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Alllyn
; APPLICANT: Olsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 454
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-454
Query Match 95.9%; Score 1803; DB 9; Length 355;
Best Local Similarity 99.2%; Pred. No. 6.1e-166; Indels 0; Gaps 0;
Matches 352; Conservative 1; Mismatches 2
Qy 2 MMSGQGKRLVMVAGGTGGHVFPGLAVAHLMAGQVQWRLGTADRMADLVPKHGIEIDFI 61
Db 1 MMSGQGKRLVMVAGGTGGHVFPGLAVAHLMAGQVQWRLGTADRMADLVPKHGIEIDFI 60
Qy 62 RISGLRGKIKALIAAPLRIFNARQARIMKAYKPDVVLGNGGYVSGPGGLAANSLGIP 121
Db 61 RISGLRGKIKALIAAPLRIFNARQARIMKAYKPDVVLGNGGYVSGPGGLAANSLGIP 120
Qy 122 VVLEHQNGIAGLTNKLARIATKVMQAEPCAFNVAEVVGNPVRTDVLALPLPQORLAGR 181
Db 121 VVLEHQNGIAGLTNKLARIATKVMQAEPCAFNVAEVVGNPVRTDVLALPLPQORLAGR 180
Qy 182 GPRVRLVVGSGQCARILNQTMPQVAAKLGDSVTIWHQSGKGSQQSVQVQYAEAGQPHKV 241
Db 181 GPRVRLVVGSGQCARILNQTMPQVAAKLGDSVTIWHQSGKGSQQSVQVQYAEAGQPHKV 240
Qy 242 TEFIDMAAYAWADVVCVCSGALTIVSEIAAAGLPALFVFPQHKDQOQYVNALPLEKAGA 301
Db 241 TEFIDMAAYAWADVVCVCSGALTIVSEIAAAGLPALFVFPQHKDQOQYVNALPLEKAGA 300
Qy 302 AKTIEQPQLSVDAVANTLAGSRETLTMAERARAASIPDATERVANEVSRVARA 356
Db 301 AKTIEQPQLSVDAVANTLAGSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 3
US-09-815-242-10031
; Sequence 10031, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chisen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10031
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10031

Query Match 95.9%; Score 1803; DB 9; Length 355;
Best Local Similarity 99.2%; Pred. No. 6,1e-166;
Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 MSGCGKRLMYMAGGTGGHVPFGLAVAHLMACQGWVRWLGTAADMEADLVPKHGIEIDFI 61
DB 1 MSGCGKRLMYMAGGTGGHVPFGLAVAHLMACQGWVRWLGTAADMEADLVPKHGIEIDFI 60

QY 62 RISGLRGKIGKALIAAPLRIFNWQARAIMKAYKPDVVLGCGVVGPGGLAASLGIP 121
DB 61 RISGLRGKIGKALIAAPLRIFNWQARAIMKAYKPDVVLGCGVVGPGGLAASLGIP 120

QY 122 VVLEHONGIAGLTNKLARIATKMQAEPGAPPAEVEVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLEHONGIAGLTNKLARIATKMQAEPGAPPAEVEVGNPVRTDVLALPLPQORLAGRE 180

QY 182 GPRVVLVVGSGGARGILNQTMPQVAAKLGDSVITWHQSGKSGQSQVEQAYAEAGOPQHKV 241
DB 181 GPRVVLVVGSGGARGILNQTMPQVAAKLGDSVITWHQSGKSGQSQVEQAYAEAGOPQHKV 240

QY 242 TEFIDMAAAYAWADVVVCRSGALTIVSEIAAAGLPALFVPPQHKDRQOQYWNALPLEKAGA 301
DB 241 TEFIDMAAAYAWADVVVCRSGALTIVSEIAAAGLPALFVPPQHKDRQOQYWNALPLEKAGA 300

QY 302 AKIIEPOLSVDAVANTLAGSRETLLTMAERARAASIPDATERVANEVSRVARA 356
DB 301 AKIIEPOLSVDAVANTLAGSRETLLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 4
US-10-282-122A-56415
; Sequence 56415, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56415
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-56415

Query Match 95.9%; Score 1803; DB 12; Length 355;
Best Local Similarity 99.2%; Pred. No. 5,1e-166;
Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 MSGCGKRLMYMAGGTGGHVPFGLAVAHLMACQGWVRWLGTAADMEADLVPKHGIEIDFI 61
DB 1 MSGCGKRLMYMAGGTGGHVPFGLAVAHLMACQGWVRWLGTAADMEADLVPKHGIEIDFI 60

QY 62 RISGLRGKIGKALIAAPLRIFNWQARAIMKAYKPDVVLGCGVVGPGGLAASLGIP 121
DB 61 RISGLRGKIGKALIAAPLRIFNWQARAIMKAYKPDVVLGCGVVGPGGLAASLGIP 120

QY 122 VVLEHONGIAGLTNKLARIATKMQAEPGAPPAEVEVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLEHONGIAGLTNKLARIATKMQAEPGAPPAEVEVGNPVRTDVLALPLPQORLAGRE 180

QY 182 GPRVVLVVGSGGARGILNQTMPQVAAKLGDSVITWHQSGKSGQSQVEQAYAEAGOPQHKV 241
DB 181 GPRVVLVVGSGGARGILNQTMPQVAAKLGDSVITWHQSGKSGQSQVEQAYAEAGOPQHKV 240

QY 242 TEFIDMAAAYAWADVVVCRSGALTIVSEIAAAGLPALFVPPQHKDRQOQYWNALPLEKAGA 301
DB 241 TEFIDMAAAYAWADVVVCRSGALTIVSEIAAAGLPALFVPPQHKDRQOQYWNALPLEKAGA 300

QY 302 AKIIEPOLSVDAVANTLAGSRETLLTMAERARAASIPDATERVANEVSRVARA 356
DB 301 AKIIEPOLSVDAVANTLAGSRETLLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 5
US-10-282-122A-75189
; Sequence 75189, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert

APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 75189
LENGTH: 355
TYPE: PRT
ORGANISM: Salmonella typhimurium
US-10-282-122A-75189

Query Match 89.7%; Score 1687; DB 12; Length 355;
Best Local Similarity 91.2%; Pred. No. 1.1e-154;
Matches 323; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

QY 2 MSGGKRLMVMAGGTGGHVFPLGLAVAHHLMAQGVQVRLGTADRMADLVPKHGIDIFI 61
DB 1 MSGGKRLMVMAGGTGGHVFPLGLAVAHHLMAQGVQVRLGTADRMADLVPKHGIDIFI 60
QY 62 RISLRGKGIKALIAAPLIRFNARQARAIWKYKPDVVLGMGGYVSGPGGLAASLGIP 121
DB 61 RISLRGKGVKALLAAPLIRFNARQARAIWKYKPDVVLGMGGYVSGPGGLAASLGIP 120
QY 122 VYLHEQNGIAGLTNKLARIATKVMQAEFGAPFNAEVVGNPVRTDVLALPLPQVRLAGRE 181
DB 121 VYLHEQNGIAGLTNKLAKIATTVMQAFPGAPFNAEVVGNPVRTDVLALPLPQVRLAGRD 180
QY 182 GPRVVLVVGSGGARILNQTWPQVAAKLGDSVIIWHQSGKGSQSVQVAYAEAGQPHKV 241
DB 181 GPRVVLVVGSGGARVLNQTWPQVAAKLGDSVIIWHQSGKGAQLTVEQVAGAGQPHKV 240
QY 242 TEFIDDDMAAAVADVVVCRSGALTVSEIAAAGLPALFVFPQHKDRQOQYNNALPLEKAGA 301
DB 241 TEFIDDDMAAAVADVVVCRSGALTVSEIAAAGLPALFVFPQHKDRQOQYNNALPLENAGA 300
QY 302 AKIIEQPQLSVDAVANTLWAGSRETLTMAERARAASIPDATERVANEVSRVAR 355
DB 301 AKIIEQPQFTVEAVADTLWAGSREALLTMAERARAVSIPDATERVANEVSRVAR 354

RESULT 6
US-09-815-242-14078
Sequence 14078, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14078
LENGTH: 355
TYPE: PRT
ORGANISM: Salmonella typhi
US-09-815-242-14078

Query Match 89.2%; Score 1677; DB 9; Length 355;
Best Local Similarity 90.7%; Pred. No. 9.9e-154;
Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;

QY 2 MSGGKRLMVMAGGTGGHVFPLGLAVAHHLMAQGVQVRLGTADRMADLVPKHGIDIFI 61
DB 1 MSGGKRLMVMAGGTGGHVFPLGLAVAHHLMAQGVQVRLGTADRMADLVPKHGIDIFI 60
QY 62 RISLRGKGIKALIAAPLIRFNARQARAIWKYKPDVVLGMGGYVSGPGGLAASLGIP 121
DB 61 RISLRGKGVKALLAAPLIRFNARQARAIWKYKPDVVLGMGGYVSGPGGLAASLGIP 120
QY 122 VYLHEQNGIAGLTNKLARIATKVMQAEFGAPFNAEVVGNPVRTDVLALPLPQVRLAGRE 181
DB 121 VYLHEQNGIAGLTNKLAKIATTVMQAFPGAPFNAEVVGNPVRTDVLALPLPQVRLAGRD 180
QY 182 GPRVVLVVGSGGARILNQTWPQVAAKLGDSVIIWHQSGKGSQSVQVAYAEAGQPHKV 241
DB 181 GPRVVLVVGSGGARVLNQTWPQVAAKLGDSVIIWHQSGKGAQLTVEQVAGAGQPHKV 240
QY 242 TEFIDDDMAAAVADVVVCRSGALTVSEIAAAGLPALFVFPQHKDRQOQYNNALPLEKAGA 301
DB 241 TEFIDDDMAAAVADVVVCRSGALTVSEIAAAGLPALFVFPQHKDRQOQYNNALPLENAGA 300
QY 302 AKIIEQPQLSVDAVANTLWAGSRETLTMAERARAASIPDATERVANEVSRVAR 355
DB 301 AKIIEQPQFTVEAVADTLWAGSREALLTMAERARAVSIPDATERVANEVSRVAR 354

RESULT 7
US-10-282-122A-76021
Sequence 76021, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangau
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John

APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR FILING DATE: 2000-05-26
PRIOR FILING DATE: 2000-09-06
PRIOR FILING DATE: 2000-09-06
PRIOR FILING DATE: 2000-10-23
PRIOR FILING DATE: 2000-11-27
PRIOR FILING DATE: 2000-12-22
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1
SEQ ID NO 76021
LENGTH: 355
TYPE: PRT
ORGANISM: Salmonella typhi
US-10-282-122A-76021

Query Match 89.2%; Score 1677; DB 12; Length 355;
Best Local Similarity 90.7%; Pred. No. 9.9e-154;
Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;

QY 2 MSGQKRLMVMAGTGGHVFGLAVAHLMQAQGVWVWLTADRMADLVPKHGIEDIFI 61
DB 1 MSGQKRLMVMAGTGGHVFGLAVAHLMQAQGVWVWLTADRMADLVPKHGIEDIFI 60
QY 62 RISGLRGKIHAKLAAPLIRIFNAWRQARAIMKAYKPDVWVLMGQYVSGPGLAASLGIP 121
DB 61 RISGLRGKIVKALLAAPLIRIFNAWRQARAIMKAYKPDVWVLMGQYVSGPGLAASLGIP 120
QY 122 VVLEHONGINGLTKWLARIATKMOAEPGAPFNAEVVGNPVRTDVLALPLPOORLAGRE 181
DB 121 VVLEHONGINGLTKWLAKIATVQAFFPGAPFNAEVVGNPVRTDVLALPLPOORLAGRD 180
QY 182 GPRVVLVVGSGGARILNQTMPQVAALGDSVVIWHQSGKGSQSVQAYAEAGQPOHKV 241
DB 181 GPRVVLVVGSGGARVNLNQTMPQVAALGDSVVIWHQSGKGAQLTVQAYAGQPOHKV 240
QY 242 TEFIDDMAAAYADVVTVCRSALTYSIAAGLPALFVPEHDKROQYNNALPLEKAGA 301
DB 241 TEFIDDMAAAYADVVTVCRSALTYSIAAGLPALFVPEHDKROQYNNALPLENAGA 300
QY 302 AKIIEQPOLSDVAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRAV 355
DB 301 AKIIEQPOLTEAVADTLAGWSREALLTMAERARAASIPDATERVANEVSRAV 354

RESULT 8
US-10-282-122A-59581
Sequence 59581, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangshu
APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR FILING DATE: 2000-05-23
PRIOR FILING DATE: 2000-05-26
PRIOR FILING DATE: 2000-05-26
PRIOR FILING DATE: 2000-09-06
PRIOR FILING DATE: 2000-09-06
PRIOR FILING DATE: 2000-09-09
PRIOR FILING DATE: 2000-10-23
PRIOR FILING DATE: 2000-11-27
PRIOR FILING DATE: 2000-12-22
PRIOR FILING DATE: 2001-02-09
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1
SEQ ID NO 59581
LENGTH: 348
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59581

Query Match 85.3%; Score 1604; DB 12; Length 348;
Best Local Similarity 88.8%; Pred. No. 1.1e-146;
Matches 308; Conservative 19; Mismatches 20; Indels 0; Gaps 0;

QY 10 MYMAGTGGHVPGLAVAHLMQAQGVWVWLTADRMADLVPKHGIEDIFIRISGLRGK 69
DB 1 MYMAGTGGHVPGLAVAHLMQAQGVWVWLTADRMADLVPKHGIEDIFIRISGLRGK 60
QY 70 GIKALITAAPLIRIFNAWRQARAIMKAYKPDVWVLMGQYVSGPGLAASLGIPVVLHEONG 129
DB 61 GIKAQALLAPVIRIFNAWRQARAIMKAYKPDVWVLMGQYVSGPGLAASLGIPVVLHEONG 120
QY 130 IAGLTNNKLIARITATKMOAEPGAPFNAEVVGNPVRTDVLALPLPOORLAGREGPVRVLV 189
DB 121 IAGLTNNKLIARITATKMOAEPGAPFNAEVVGNPVRTDVLALPLPOORLAGREGPVRVLV 180
QY 190 GSGQGARILNQTMPQVAALGDSVVIWHQSGKGSQSVQAYAEAGQPOHKVTEFIDDMA 249
DB 181 GSGQGARVNLNQTMPQVAALGDSVVIWHQSGKGSQSVQAYAEAGQPOHKVTEFIDDMA 240
QY 250 AAYAWADVVCRSALTYSIAAGLPALFVPEHDKROQYNNALPLEKAGAAKILEQPC 309
DB 241 AAYAWADVVCRSALTYSIAAGLPALFVPEHDKROQYNNALPLEKAGAAKILEQPC 300
QY 310 LSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRAV 356
DB 301 FTVEAVASTLASWDRETLTMAERARAASIPDATERVANEVSRAV 347

RESULT 9

US-10-282-122A-73040
; Sequence 73040, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 73040
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Salmonella paratyphi A
; NAME/KEY: MISC FEATURE
; LOCATION: (259)..(259)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (271)..(271)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-73040

Query Match 82.4%; Score 1549.5; DB 12; Length 348;
Best Local Similarity 86.5%; Pred. No. 2.2e-141;
Matches 302; Conservative 16; Mismatches 30; Indels 1; Gaps 1;
QY 2 MSGGKELMWAGTGGHVPFGLAVAHLMMAQGWVRLGTADRMADLVPRKHGIDIFI 61
DB 1 MSGGKELMWAGTGGHVPFGLAVAHLMMAQGWVRLGTADRMADLVPRKHGIDIFI 60
QY 62 RISGLRGKIGKALIAAPLRFNARQARAIMKAYKDDVVLGMGGYVSGPGGLAANSLGIP 121
DB 61 RISGLRGKIGKALIAAPLRFNARQARAIMKAYKDDVVLGMGGYVSGPGGLAANSLGIP 120
QY 122 VILHEQNGIAGLTNKNLARIATKMQAEPGAFNAEVVGNFVTDVLAFLPQORLAGRE 181
DB 121 VILHEQNGIAGLTNKNLARIATKMQAEPGAFNAEVVGNFVTDVLAFLPQORLAGRD 180

QY 182 GPRVVLVWGSQGARILNQTMPQVAALGDSVLIWHQSGKGSQSVQAYARAGQPQHKV 241
DB 181 GPRVVLVWGSQGARVNLNQTMPQVAARLGDVTIWHQSGKGAQLTVEQYAGTGPQHKV 240
QY 242 TEFIDMAAAYAWADVVVCRSGNLTYSIATAAGLPAFLVFPQHKDSQQYWNALPLEKAGA 301
DB 241 TEFIDMAAAYAWADVVVXVSGALTVSEIAXPCLPALFVFPQHKDRQQYWNALPLENAGA 300
QY 302 AKIIEQPQLSVDVANTLAGWSRETLLTMAERARAASIPDATERVANEV 350
DB 301 AKIIEQPQFTVDVSPNT-GGVVAEALLTMAEVPAAVSIPTATERVASEV 348

RESULT 10
US-10-282-122A-77926
; Sequence 77926, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 73040
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Salmonella paratyphi A
; NAME/KEY: MISC FEATURE
; LOCATION: (259)..(259)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (271)..(271)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-73040

QY 2 MSGGKELMWAGTGGHVPFGLAVAHLMMAQGWVRLGTADRMADLVPRKHGIDIFI 61
DB 1 MSGGKELMWAGTGGHVPFGLAVAHLMMAQGWVRLGTADRMADLVPRKHGIDIFI 60
QY 62 RISGLRGKIGKALIAAPLRFNARQARAIMKAYKDDVVLGMGGYVSGPGGLAANSLGIP 121
DB 61 RISGLRGKIGKALIAAPLRFNARQARAIMKAYKDDVVLGMGGYVSGPGGLAANSLGIP 120
QY 122 VILHEQNGIAGLTNKNLARIATKMQAEPGAFNAEVVGNFVTDVLAFLPQORLAGRE 181
DB 121 VILHEQNGIAGLTNKNLARIATKMQAEPGAFNAEVVGNFVTDVLAFLPQORLAGRD 180

QY 182 GPRVVLVWGSQGARILNQTMPQVAALGDSVLIWHQSGKGSQSVQAYARAGQPQHKV 241
DB 181 GPRVVLVWGSQGARVNLNQTMPQVAARLGDVTIWHQSGKGAQLTVEQYAGTGPQHKV 240
QY 242 TEFIDMAAAYAWADVVVCRSGNLTYSIATAAGLPAFLVFPQHKDSQQYWNALPLEKAGA 301
DB 241 TEFIDMAAAYAWADVVVXVSGALTVSEIAXPCLPALFVFPQHKDRQQYWNALPLENAGA 300
QY 302 AKIIEQPQLSVDVANTLAGWSRETLLTMAERARAASIPDATERVANEV 350
DB 301 AKIIEQPQFTVDVSPNT-GGVVAEALLTMAEVPAAVSIPTATERVASEV 348

RESULT 10
US-10-282-122A-77926
; Sequence 77926, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 77926
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-77926

Query Match 78.2%; Score 1471; DB 12; Length 356;
Best Local Similarity 78.3%; Pred. No. 9.1e-134;
Matches 278; Conservative 36; Mismatches 41; Indels 0; Gaps 0;
QY 2 MSGGKELMWAGTGGHVPFGLAVAHLMMAQGWVRLGTADRMADLVPRKHGIDIFI 61
DB 1 MSGGKELMWAGTGGHVPFGLAVAHLMMAQGWVRLGTADRMADLVPRKHGIDIFI 60
QY 62 RISGLRGKIGKALIAAPLRFNARQARAIMKAYKDDVVLGMGGYVSGPGGLAANSLGIP 121

Db 61 KISGLRGKGLMAQLTAPIRIYRAVROAQKIMRDYQPNVLMGMYVSGPGGLAAMLGVP 120
QY 122 VVLEHQNGIAGLTNKLARATATKMOAEPGAFNAEVGNVPTDVLALPLPQORLAGRE 181
Db 121 VVLEHQNGIAGLTNKLARATATKMOAEPGAFNAEVGNVPTDVLALPLPQORLAGRE 180
QY 182 GPVRVLVVGSGQAGRIINQTMPTQVAAKLGDSVLIHQSGKGSQSQSVQEQAYAGAPQHKV 241
Db 181 GPVRVLVVGSGQAGRIINQTMPTQVAAKLGDSVLIHQSGKGSQSQSVQEQAYAGAPQHKV 240
QY 242 TEFDIDMAAAYAWADVVCBSGALTVSEIAAAGLPAIFVFPQHKDRQOQYWNALPLEKAGA 301
Db 241 VEFIDMAAAYAWADVVCBSGALTVSEIAAAGLPAIFVFPQHKDRQOQYWNALPLEKAGA 300
QY 302 AKIEQPOLSDVAVANTLACWSRETLTWAERARAASIPDATERVANEYSRVAR 356
Db 301 AKIEQPOLSDVAVANTLACWSRETLTWAERARAASIPDATERVANEYSRVAR 355

RESULT 11

US-10-282-122A-68811
; Sequence 68811, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 68811
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-10-282-122A-68811

Query Match 71.5%; Score 1345.5; DB 12; Length 360;
Best Local Similarity 72.1%; Pred. No. 1.4e-121;
Matches 259; Conservative 40; Mismatches 55; Indels 5; Gaps 1;

QY 2 MSGGKRLMVMAGTGGHVPFGLAVAHHLMAQGWVWILGTADNMEADLYPKHGIEIDFI 61
Db 1 MSERKRLMVMAGTGGHVPFGLAVAHHLMAQGWVWILGTADNMEADLYPKHGIEIYI 60
QY 62 RISGLRGKIKALIAAPLRIFNAWROARAIAWKYKPDVWLCMGYVSGPGGLAASLGIP 121
Db 61 RISGLRGKIKALIAAPLRIFNAWROARAIAWKYKPDVWLCMGYVSGPGGLAASLGIP 120
QY 122 VVLEHQNGIAGLTNKLARATATKMOAEPGAFNAEVGNVPTDVLALPLPQORLAGRE 181
Db 121 VVLEHQNGIAGLTNKLARATATKMOAEPGAFNAEVGNVPTDVLALPLPQORLAGRE 180
QY 182 GPVRVLVVGSGQAGRIINQTMPTQVAAKLGDSVLIHQSGKGSQSQSVQEQAY 236
Db 181 GPVRVLVVGSGQAGRIINQTMPTQVAAKLGDSVLIHQSGKGSQSQSVQEQAY 240
QY 237 POHKVTEFDIDMAAAYAWADVVCBSGALTVSEIAAAGLPAIFVFPQHKDRQOQYWNALPL 296
Db 241 SEYKTEFDIDMAAAYAWADVVCBSGALTVSEIAAAGLPAIFVFPQHKDRQOQYWNALPL 300
QY 297 EKAGAAKIEQPOLSDVAVANTLACWSRETLTWAERARAASIPDATERVANEYSRVAR 355
Db 301 ENAGAARIIEQNDLTPEATDLENWDRHQLMLMAEKAQSVAITDATERVANVIEVAK 359

RESULT 12

US-10-282-122A-55838
; Sequence 55838, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 55838
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Enterobacter cloacae

US-10-282-122A-55838

```

Query Match      70.3%; Score 1323; DB 12; Length 281;
Best Local Similarity 91.3%; Pred. No. 1.4e-119;
Matches 253; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 5 QKRLMWAGGTGGHVFPGCLAVAHLMAGQWVRWLGTDADMEADLVPKHGIEIDFIRIS 64
DB 3 QKRLMWAGGTGGHVFPGCLAVAHLMAGQWVRWLGTDADMEADLVPKHGIEIDFIRIS 62
QY 65 GLRGKIGIKALIAAPLRIFNARQARAIMKAYKPDVVLGMGGVSGPGGLAAWSLGI PVVL 124
DB 63 GLRGKGLKAMLLAPVRIFNARQARAIMKAYKPDVVLGMGGVSGPGGLAAWSLGI PVVL 122
QY 125 HQNGIAGLTNKLARIATKMAEPGAPPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 184
DB 123 HQNGIAGLTNKLARIATKMAEPGAPPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 182
QY 185 RVLVVGSGGAGILNQTMPQVAAKLGDSVIIHQSGKGSQQSVQAYAEAGQPQHKVTEF 244
DB 183 RVLVVGSGGAGILNQTMPQVAAKLGDSVIIHQSGKGSQQSVQAYAEAGQPQHKVTEF 242
QY 245 IDDMAAYAWADVVCRSGLTIVSEIAAAGLPALFVP 281
DB 243 IDDMAAYAWADVVCRSGLTIVSEIAAAGLPALFVP 279

```

RESULT 13

```

US-10-282-122A-77456
; Sequence 77456, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77456
; LENGTH: 354

```

```

; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-77456

```

```

Query Match      61.4%; Score 1154; DB 12; Length 354;
Best Local Similarity 64.2%; Pred. No. 4.8e-103;
Matches 224; Conservative 48; Mismatches 77; Indels 0; Gaps 0;

QY 2 MSQGRKLMVAGGTGGHVFPGCLAVAHLMAGQWVRWLGTDADMEADLVPKHGIEIDFIRIS 61
DB 1 MNKNKKLMVAGGTGGHVFPGCLAVAHLMAGQWVRWLGTDADMEADLVPKHGIEIDFIRIS 60
QY 62 RISGLRGKIGIKALIAAPLRIFNARQARAIMKAYKPDVVLGMGGVSGPGGLAAWSLGI PVVL 121
DB 61 QVKGRLGQGLMRLKAPFQVNVAILQARRHLLTYQPDVVLGMGGVSGPGGLAAWSLGI PVVL 120
QY 122 VYLHQNGIAGLTNKLARIATKMAEPGAPPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 181
DB 121 VYLHQNGIAGLTNKLARIATKMAEPGAPPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 180
QY 182 GPVVRVLVVGSGGAGILNQTMPQVAAKLGDSVIIHQSGKGSQQSVQAYAEAGQPQHKV 241
DB 181 GAIIRLVVGSGGAGILNQTMPQVAAKLGDSVIIHQSGKGSQQSVQAYAEAGQPQHKV 240
QY 242 TEFIDDMAAYAWADVVCRSGLTIVSEIAAAGLPALFVPFQHKDRCQYWNALPLEKAGA 301
DB 241 TEFIDDMAAYAWADVVCRSGLTIVSEIAAAGLPALFVPFQHKDRCQYWNALPLEKAGA 300
QY 302 AKTIEQPOLSVDAVANTLWAGSRETIITWAEARARASIPDATEVANEV 350
DB 301 AKTIEQPOLSVDAVANTLWAGSRETIITWAEARARASIPDATEVANEV 349

```

RESULT 14

```

US-10-282-122A-66899
; Sequence 66899, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77456
; LENGTH: 354

```

```

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0

```

PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 58422
LENGTH: 351
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-10-282-122A-58422

Query Match 56.4%; Score 1064.5; DB 12; Length 351;
Best Local Similarity 61.5%; Pred. No. 2.2e-94;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;
QY 5 QGKELVMAGTGGHVPFGLAVAHLMACQVRLGTADRMEDLVPEKHGIEIDFIRIS 64
DB 2 KKKLLVMAGTGGHVPFPAVAQTLOKQEWDCWLGTKDQRMQAQLVPEKIGIPVIL 61
QY 65 GLRGKIGIKALIAAPLRIENAWRQARAIMKAYKEDVVLGMGGYVSGPGGLAANSLGIPVIL 124
DB 62 GLRGKIGIKALLNAPFAIFRAVLQAKKIIQEEKEDAVLGMGGYVSGPAGVAALKCGVPIIL 121
QY 125 HEQNGIAGLTKNLARIATKVMQAEPCAFNAEVVGNPVRTDVLALPQORLAGREGPV 184
DB 122 HEQNAIAGLTKNLGKIATCVLQAFPTAFPAEAVVGNPVRDLFEMPNDPIRFSDBEKL 181
QY 185 RVLVVGSGQARIINOTMPOVAAKLGDVSIIVHQSQSGQSQVQEAAGPQHKVITF 244
DB 182 RVLVVGSGQARVNLHTLPKVVAQLADKLEFPHQVKGAVEEVSQLYGE-NLEQVKITEF 240
QY 245 IDMAAAYAVADVVVCRSGALTVSEIAAAGLPALFVFPQHKDQYVWALPLEKAGAAKI 304
DB 241 IDNMAEAYAVADVVVCRSGALTVCEIAAVGAAAFVFPQHKDQYVWALPLEKAGAAKI 300
QY 305 IEQPLSDVADVANTLAGMSRETLITMAERARASIPDATERVANEVSR 352
DB 301 IEQADLTPEILVNLTKNLTRENLLOMALKAKTMSMFAAORVAEVIQ 348

RESULT 17

US-09-829-275-2
Sequence 275-2, Application US/09829275
Publication No. US20030077803A1
GENERAL INFORMATION:
APPLICANT: WALKER, SUZANNE
TITLE OF INVENTION: CRYSTALS OF THE ESCHERICHIA COLI MEMBRANE-ASSOCIATED
TITLE OF INVENTION: GLYCOSYLTRANSFERASE (MURG) PROTEIN, ATOMIC COORDINATES
TITLE OF INVENTION: AND THREE DIMENSIONAL STRUCTURES THEREOF, ATOMIC
TITLE OF INVENTION: COORDINATES AND THREE DIMENSIONAL STRUCTURES OF
TITLE OF INVENTION: BINDING DOMAINS THEREOF, IMAGES THEREOF, AND METHODS
TITLE OF INVENTION: OF CRYSTALLIZING MURG PROTEINS, MODELS OF UDP-
TITLE OF INVENTION: GLYCOSYLTRANSFERASES, MURG PROTEINS AND BINDING SITES,
TITLE OF INVENTION: METHODS OF MAKING MODELS, METHODS OF USING MODELS OF
TITLE OF INVENTION: MURG, COMPOUNDS THAT BIND, INHIBIT OR STIMULATE MURG
TITLE OF INVENTION: PROTEINS, AND THERAPEUTIC COMPOSITIONS THEREOF
FILE REFERENCE: 4555-105
CURRENT APPLICATION NUMBER: US/09/829,275
CURRENT FILING DATE: 2001-09-11
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 351
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-829-275-2

Query Match 56.4%; Score 1061.5; DB 10; Length 351;
Best Local Similarity 61.5%; Pred. No. 4.3e-94;
Matches 214; Conservative 43; Mismatches 90; Indels 1; Gaps 1;
QY 5 QGKELVMAGTGGHVPFGLAVAHLMACQVRLGTADRMEDLVPEKHGIEIDFIRIS 64

DB 2 KKKLLVMAGTGGHVPFPAVAQTLOKQEWDCWLGTKDQRMQAQLVPEKIGIPVIL 61
QY 65 GLRGKIGIKALIAAPLRIENAWRQARAIMKAYKEDVVLGMGGYVSGPGGLAANSLGIPVIL 124
DB 62 GLRGKIGIKALLNAPFAIFRAVLQAKKIIQEEKEDAVLGMGGYVSGPAGVAALKCGVPIIL 121
QY 125 HEQNGIAGLTKNLARIATKVMQAEPCAFNAEVVGNPVRTDVLALPQORLAGREGPV 184
DB 122 HEQNAIAGLTKNLGKIATCVLQAFPTAFPAEAVVGNPVRDLFEMPNDPIRFSDBEKL 181
QY 185 RVLVVGSGQARIINOTMPOVAAKLGDVSIIVHQSQSGQSQVQEAAGPQHKVITF 244
DB 182 RVLVVGSGQARVNLHTLPKVVAQLADKLEFPHQVKGAVEEVSQLYGE-NLEQVKITEF 240
QY 245 IDMAAAYAVADVVVCRSGALTVSEIAAAGLPALFVFPQHKDQYVWALPLEKAGAAKI 304
DB 241 IDNMAEAYAVADVVVCRSGALTVCEIAAVGAAAFVFPQHKDQYVWALPLEKAGAAKI 300
QY 305 IEQPLSDVADVANTLAGMSRETLITMAERARASIPDATERVANEVSR 352
DB 301 IEQADLTPEILVNLTKNLTRENLLOMALKAKTMSMFAAORVAEVIQ 348

RESULT 18

US-10-282-122A-50225
Sequence 50225, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Chlesen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 50225
LENGTH: 367
TYPE: PRT
ORGANISM: Burkholderia mallei
US-10-282-122A-50225

Query Match	46.3%;	Score 871;	DB 12;	Length 367;
Best Local Similarity	50.7%;	Pred. No. 1.3e-75;		
Matches 183;	Conservative 57;	Mismatches 115;	Indels 6;	Gaps 3;
QY	2	MSGQKRLMVMAGGTGGHVFPGLAVAHLMQGMQVREWLGTADRMEDALVPKHGIEIDFI	61
Db	1	MTSTORTLMVMAGGTGGHVFPGLAVAHMQAQGRVVWLGNPAGMEATLVRHGIPMEV	60
QY	62	RISGLRGKGIKALIAAPLRIFNARQARALIMKAYKDDVLGMGQYVSGPGLAANSGLIP	121
Db	61	RFGGLRGKGLATKALPFNLRLACAQSLRALRKVPDWLGMGGYITFPAGLTVTLTGRP	120
QY	122	VYLHEQNGIAGLTNKLARLAKVMQEPGCAFFNAEVVGNPVRTDLVALPLPQQLAGRE	181
Db	121	LVLHEQNSIAGLTNKLAKLAKRVLVAFPGLPNAEWGTFNTELTARTEPPQARVAARS	180
QY	182	GPVRVLVVGSGQGARILNCTMQVAANKLG--DSVITMQSGKSGSQSVQCAVAEAG--Q	236
Db	181	GKRLVLVVGSGLGAALNEVVPRALALLADPERQVVHQAGAKHIDTLKENTAEAGLSG	240
QY	237	POHKYTEPIDDMAAYAWADVVVCRSGALTVSRISAAAGLPALFVFPQHK-DRQQYNNALP	295
Db	241	SDVALVPFIDDMAGAYANADLIVICRSGAMTVABIIAAGVAAALFVPPPHAVDDHQTNAEF	300
QY	296	LEKAGAAKIIEQPOLSDVDVAANTLAGHSRETLITMAERAPAAASIPDATERVANEYSRVAR	355
Db	301	LAEGQAANLVQORDLSAELLADWLRCGQRSRLAAMAEERSLAKPATDEVARVCAAVAG	360
QY	356	A	356	A 356
Db	361	A	361	A 361

RESULT 19
US-10-282-122A-49098
; Sequence 49098, Application US/10282122A
; Publication NO. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09

```

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49098
; LENGTH: 372
; TYPE: FRT
; ORGANISM: Burkholderia fungorum
; US-10-282-122A-49058

Query Match          45.1%; Score 849; DB 12; Length 372;
Best Local Similarity 49.9%; Pred. No. 1.9e-73;
Matches 178; Conservative 60; Mismatches 113; Indels 6; Gaps 3;

QY      7 KRLVMVAGGTGGHVFPGLVAHHLMAQGHQVWGLFADRMEDLVPKHGIDIRISGL 66
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      7 RTLVMVAGGTGGHVFPGLVAHHLMAQGHQVWGLFADRMEDLVPKHGIPMEYVRFGL 66
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      67 RGKGIKALIAPIRINAFNRQARIMKAYKPDVWLGMGYVSGPGLAAMSLGIPVVLHE 126
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      67 RGKGLKTKLMLPLNLLRACTQSUSLVLRYPKPDVWLGMYITFPAGLTALSGRPLVHE 126
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      127 QNGIAGLTNKLARIATKVMQAEPPGAFNAEVVGNPRTDVIALPLPORLAGREGPVVR 186
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      127 QNSIAGLANKLAKVAKRVLVAPFPNALPHCEWTGNPIRAELAGALAPKARYAORSGLNV 186
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      187 LVVGSOGARILNQTWPQVAKL--GDSVI IWHOSKGSQQSQVEQAYAEAG--QPQHKV 241
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      187 LVVGSJGAALNEVVPRAVALLAPNERPRIVHQAGAKHIEALREYAAAGLQAGADVEL 246
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      242 TEFIDDDAAAAYANADVVCVRSRGALTYSIEIAAAGLPALEVPFQHK--DROQYNALPLEKAG 300
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      247 VPFIDDMTSAYANADLVICRSGAMTVSEISAVGVAALFVPPPYAVDDHQTNAAFADNG 306
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      301 AAKIEQPQLSDVAVANTLACNSRETLTWAERARAASIPDATERVANEVSRAAL 357
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      307 AALVVQQRDLSAETLADMLSSQRTRETLAEVRSRLAKPDQATEQVQATCATVAGSI 363
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 20
US-10-282-122A-51374
; Sequence 51374, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

RESULT 20

US-10-282

; Sequenc

; Publica

; GENERAL

; APPLIC

APPLIC.

APPLIC.

APPLIC.

APPLICABLE

APPLICABLE

APPLICABILITY

ADDITIONAL

APPLT

APPLIC

APPLIC

TITLE

FILE R

;
CURRENT

;

;
PRIOR

;
PRIOR

;
PRIOR
PRIOR

; PRIOR PRIOR

PRIOR
PRIOR

PRISON
PRISON

PRISON
PRISON

PRTOR
PRTOR

✓ : PRIOR

PRIOR

PRIOR

10

1000 2000

100 500 1000 1500 2000

Search completed: June 7, 2004, 07:20:39
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 07:01:50 ; Search time 21 Seconds
(without alignments)

1667.320 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 MMSGQKRLVMAGTGGVH.....RVANEVSRLVAREHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR.78.*

1: Piri.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
1	1803	95.9	355	1	BVECMG	UDP-N-acetylglucos
2	1793	95.3	355	2	F90640	hypothetical prote
3	1793	95.3	355	2	F85491	hypothetical prote
4	1677	89.2	355	2	AE0518	UDP-N-acetylglucos
5	1471	78.2	356	2	AH0068	UDP-N-acetylglucos
6	1154	61.4	354	2	B82081	UDP-N-acetylglucos
7	1054.5	56.6	351	2	D64185	UDP-N-acetylglucos
8	939	49.9	354	2	D84935	hypothetical prote
9	780	41.5	367	2	D82763	UDP-N-acetylglucos
10	779.5	41.4	355	2	A81201	UDP-N-acetylglucos
11	772.5	41.1	355	2	B81777	UDP-N-acetylglucos
12	719.5	38.3	357	2	E83094	UDP-N-acetylglucos
13	510.5	27.1	407	2	D87023	hypothetical prote
14	506	26.9	410	2	E70579	probable murG prot
15	501.5	26.7	380	2	AE3324	UDP-N-acetylglucos
16	484	25.7	361	2	F87565	hypothetical prote
17	470	25.0	363	2	E83970	UDP-N-acetylglucos
18	450.5	24.0	364	2	T34954	probable UDP-N-ace
19	442	23.5	378	2	AG2833	hypothetical prote
20	442	23.5	378	2	C97611	hypothetical prote
21	435	21.1	357	2	E86823	peptidoglycan synt
22	422	22.4	363	2	JCL1275	phospho-N-acetylm
23	422	22.4	363	2	AC1339	peptidoglycan synt
24	419	22.3	418	2	G75496	UDP-N-acetylglucos
25	418	22.2	371	2	AC1700	peptidoglycan synt
26	398.5	21.2	371	2	S76863	hypothetical prote
27	340	18.1	376	2	B97770	hypothetical prote
28	334.5	17.8	344	2	C70401	phospho-N-acetylm
29	332	17.7	431	2	F96784	hypothetical prote

30	328.5	17.5	359	2	A97175	undecaprenyl-PP-Mu
31	321.5	17.1	339	2	B72402	UDP-N-acetylglucos
32	319.5	17.0	385	2	C71699	murG protein (murG
33	315.5	16.8	384	2	A71316	probable UDP-N-glu
34	313	16.6	352	2	H71474	probable peptidogl
35	303	16.1	353	2	A81737	UDP-N-acetylglucos
36	301.5	16.0	261	2	AD1866	phospho-N-acetylm
37	284.5	15.1	357	2	F86603	peptidoglycan tran
38	284.5	15.1	357	2	C72022	UDP-N-acetylglucos
39	277	14.7	353	2	G71852	udp-n-acetylglucos
40	275	14.6	353	2	C84664	transferase, pepti
41	273	14.5	356	2	B89919	hypothetical prote
42	272.5	14.5	352	2	D79747	UDP-N-acetylglucos
43	269.5	14.3	352	2	A95080	hypothetical prote
44	263	14.0	342	2	D81306	probable UDP-N-ace
45	259.5	13.8	363	2	F70195	UDP-N-acetylglucos
46	162.5	8.6	388	2	B70878	probable transfera
47	136	7.2	392	2	S72936	UDP-glucuronosyltr
48	135.5	7.2	427	2	B95936	probable glycosylt
49	131.5	7.0	346	2	H95397	protein imported
50	131	7.0	525	2	T10478	probable 1,2-diacy
51	130.5	6.9	433	2	S77340	hypothetical prote
52	127	6.8	382	2	C59935	cell wall synthesi
53	127	6.8	519	2	H69034	conserved hypothet
54	122.5	6.5	237	2	A80807	hypothetical prote
55	122.5	6.5	559	2	F83283	precorrin-3 methyl
56	122	6.5	383	2	AG2894	conserved hypothet
57	122	6.5	388	2	B76770	hypothetical prote
58	120	6.4	411	2	F75439	probable cell wall
59	119.5	6.4	370	2	B82751	lipopolysaccharide
60	116.5	6.2	360	2	F95933	probable glycosylt
61	116.5	6.2	435	2	H86924	probable glycosylt
62	116.5	6.2	438	2	E86924	probable glycosyl
63	116	6.2	464	2	C84499	probable monogalac
64	115.5	6.1	428	2	C70670	probable glycosyl-
65	115.5	6.1	648	1	P3BPF6	p3 protein - phage
66	115	6.1	388	2	E72354	probable hexosyltr
67	115	6.1	428	2	S70670	3-deoxy-D-manno-2-
68	115	6.1	449	2	G70670	hypothetical prote
69	115	6.1	1273	2	E72611	probable Alf-depen
70	114	6.1	814	2	S65083	5-methyltetrahydro
71	113	6.0	533	2	T05092	probable 1,2-diacy
72	110.5	5.9	727	1	S17854	NADH2 dehydrogenas
73	109.5	5.8	406	2	A82455	hypothetical prote
74	109	5.8	392	1	C59851	macrolide glycosyl
75	107.5	5.7	369	2	H75345	hypothetical prote
76	106.5	5.7	360	2	E70892	hypothetical prote
77	106.5	5.7	510	2	G72464	hypothetical prote
78	105.5	5.6	338	2	C70553	hypothetical prote
79	105.5	5.6	464	2	F72512	hypothetical prote
80	105.5	5.6	1111	2	T01239	hypothetical prote
81	105	5.6	402	2	A87492	glycosyl transfera
82	105	5.6	458	2	A75386	conserved hypothet
83	105	5.6	654	2	T34613	NADH2 dehydrogenas
84	104.5	5.6	537	2	T28683	hypothetical prote
85	104	5.5	378	2	H70548	hypothetical prote
86	104	5.5	382	2	T46519	probable glycosyl
87	103.5	5.5	401	2	T45222	heat shock protein
88	103	5.5	432	2	T35527	hypothetical prote
89	103	5.5	1275	2	T18556	O-antigen biosynth
90	102.5	5.4	395	2	H72654	hypothetical prote
91	102.5	5.4	395	2	B95938	hypothetical prote
92	102.5	5.4	404	2	T35421	probable regulator
93	102	5.4	485	2	T35571	hypothetical prote
94	102	5.4	510	2	T12970	proline dehydrogen
95	102	5.4	1016	2	T31343	probable dUDP-rham
96	101	5.4	311	2	H87002	conserved hypothet
97	101	5.4	380	2	T35953	polyketide synthas
98	101	5.3	1616	2	G70668	hypothetical prote
99	100.5	5.3	370	2	T36352	hypothetical prote
100	100.5	5.3	534	2	D71344	hypothetical prote

ALIGNMENTS

RESULT 1

EVCEMG
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)pyrophosphoryl-undecaprenol N-ace
C/Species: Escherichia coli
C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 01-Mar-2002
C/Accession: JQ0544; JH0093; S40600; B64731
R/Keda, M.; Wachi, M.; Jung, H.K.; Ishino, F.; Matsuhashi, M.
Nucleic Acids Res. 18, 4014, 1990
A/Title: Nucleotide sequence involving murG and murC in the mra gene cluster region of E
A/Reference number: JQ0544; MUID: 90326550; PMID: 2197603
A/Accession: JQ0544
A/Molecule type: DNA
A/Residues: 1-355 <IKS>
A/Cross-references: EMBL:X52644; NID:942053; PIDN:CAA36867.1; PID:942055
R/Mengin-Lecreux, D.; Texier, L.; van Heijenoort, J.
Nucleic Acids Res. 18, 2810, 1990
A/Title: Nucleotide sequence of the cell-envelope murG gene of Escherichia coli.
A/Reference number: JH0093; MUID: 90251461; PMID: 2187180
A/Accession: JH0093
A/Molecule type: DNA
A/Residues: 1-355 <MEN>
A/Cross-references: EMBL:X52540; NID:942051; PIDN:CAA36776.1; PID:942052
A/Experimental source: strain K12
A/Note: it is uncertain whether Met-1 or Met-9 is the initiator
A/Note: the codon GTA given for residue 274 is inconsistent with the authors' translatio
R/Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobu
submitted to the EMBL Data Library, December 1992
A/Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2
A/Reference number: S40531
A/Accession: S40600
A/Molecule type: DNA
A/Residues: 1-355 <YUR>
A/Cross-references: EMBL:D10483; NID:9216434; PIDN:BA01355.1; PID:9216504
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shaio, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID: 97426617; PMID: 9278503
A/Accession: B64731
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-355 <BLAT>
A/Cross-references: GB:A600118; GB:U00036; NID:91786262; PIDN:AAC73201.1; PID:91786278
A/Experimental source: strain K-12, substrain MG1655
C/Genetics:
A/Gene: murG
A/Map position: 2 min
A/Function:
A/Description: involved in murein or cell envelope biosynthesis
A/Pathway: peptidoglycan biosynthesis
C/Superfamily: murG protein
C/Keywords: cell division; cell wall; glycosyltransferase; hexosyltransferase; peptidogl

Query Match 95.9%; Score 1803; DB 1; Length 355;

Best Local Similarity 99.2%; Pred. No. 1.4e-125;

Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 2 MSGQKRLVMVAGTGGHVFPGGLAVAHLMAGQWQVRLGTADRMEADLVPKHGIEIDFI 61

Db 1 MSGQKRLVMVAGTGGHVFPGGLAVAHLMAGQWQVRLGTADRMEADLVPKHGIEIDFI 60

Qy 62 RISGLRGKIGIKALIAAPLRIFNAMRQARAIKAYKPDVVLGMGGYVSGPGLAAMSLGIP 121

Db 61 RISGLRGKIGIKALIAAPLRIFNAMRQARAIKAYKPDVVLGMGGYVSGPGLAAMSLGIP 120

Qy 122 VVLEHQNGIAGLTNKLAKIATKVMQAPFGAPPNAEVVGNPVRTDVLALPLPQORLAGRE 181

Db 121 VVLEHQNGIAGLTNKLAKIATKVMQAPFGAPPNAEVVGNPVRTDVLALPLPQORLAGRE 180

Qy 182 GPRVVLVVGSGQGARILNQTWPQVAAKLGDSVTIIHQSGKGSQQSVQYAEAGQPQHKV 241

Db 242 TEFIDDMAAAAYAVADVVCRSALTIVSEIAAAGLPALFVFPQHKDRQOYWNALPLEKAGA 301

Db 241 TEFIDDMAAAAYAVADVVCRSALTIVSEIAAAGLPALFVFPQHKDRQOYWNALPLEKAGA 300

Qy 302 AKTIEQPQLSVDVAVANTLAGSRETLTMAERARAASIPDATERVANEVSVARA 356

Db 301 AKTIEQPQLSVDVAVANTLAGSRETLTMAERARAASIPDATERVANEVSVARA 355

RESULT 3

F85491

hypothetical protein murG [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: F85491

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dinalanta, E.; Potamoculis, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: A85480; MUID: 21074935; PMID: 11206551

Db 181 GPRVVLVVGSGQGARILNQTWPQVAAKLGDSVTIIHQSGKGSQQSVQYAEAGQPQHKV 240

Qy 242 TEFIDDMAAAAYAVADVVCRSALTIVSEIAAAGLPALFVFPQHKDRQOYWNALPLEKAGA 301

Db 241 TEFIDDMAAAAYAVADVVCRSALTIVSEIAAAGLPALFVFPQHKDRQOYWNALPLEKAGA 300

Qy 302 AKTIEQPQLSVDVAVANTLAGSRETLTMAERARAASIPDATERVANEVSVARA 356

Db 301 AKTIEQPQLSVDVAVANTLAGSRETLTMAERARAASIPDATERVANEVSVARA 355

RESULT 2

F90640
hypothetical protein ECs0094 [imported] - Escherichia coli (strain O157:H7, substrain RN
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C/Accession: F90640
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Gasaawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genon
A/Reference number: A99629; MUID: 21156231; PMID: 11258796
A/Accession: F90640
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-355 <HAY>
A/Cross-references: GB:BA000007; PIDN:BA833517.1; PID:813359550; GSPDB:GN00154
A/Experimental source: strain O157:H7, substrain RIMD 0509952
C/Genetics:
A/Gene: ECs0094
C/Superfamily: murG protein

Query Match 95.3%; Score 1793; DB 2; Length 355;

Best Local Similarity 98.6%; Pred. No. 7.4e-125;

Matches 350; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 MSGQKRLVMVAGTGGHVFPGGLAVAHLMAGQWQVRLGTADRMEADLVPKHGIEIDFI 61

Db 1 MSGQKRLVMVAGTGGHVFPGGLAVAHLMAGQWQVRLGTADRMEADLVPKHGIEIDFI 60

Qy 62 RISGLRGKIGIKALIAAPLRIFNAMRQARAIKAYKPDVVLGMGGYVSGPGLAAMSLGIP 121

Db 61 RISGLRGKIGIKALIAAPLRIFNAMRQARAIKAYKPDVVLGMGGYVSGPGLAAMSLGIP 120

Qy 122 VVLEHQNGIAGLTNKLAKIATKVMQAPFGAPPNAEVVGNPVRTDVLALPLPQORLAGRE 181

Db 121 VVLEHQNGIAGLTNKLAKIATKVMQAPFGAPPNAEVVGNPVRTDVLALPLPQORLAGRE 180

Qy 182 GPRVVLVVGSGQGARILNQTWPQVAAKLGDSVTIIHQSGKGSQQSVQYAEAGQPQHKV 241

Db 181 GPRVVLVVGSGQGARILNQTWPQVAAKLGDSVTIIHQSGKGSQQSVQYAEAGQPQHKV 240

Qy 242 TEFIDDMAAAAYAVADVVCRSALTIVSEIAAAGLPALFVFPQHKDRQOYWNALPLEKAGA 301

Db 241 TEFIDDMAAAAYAVADVVCRSALTIVSEIAAAGLPALFVFPQHKDRQOYWNALPLEKAGA 300

Qy 302 AKTIEQPQLSVDVAVANTLAGSRETLTMAERARAASIPDATERVANEVSVARA 356

Db 301 AKTIEQPQLSVDVAVANTLAGSRETLTMAERARAASIPDATERVANEVSVARA 355

A:Accession: P85491
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <STO>
A:Cross-references: GB:AE005174; NID:g12512793; PIDN:AAG54394.1; GSPDB:GN00145; UWGP:201
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 95.3%; Score:1793; DB 2; Length 355;
Best Local Similarity 98.6%; Pred. No. 7.4e-125;
Matches 350; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 MSGQKRLMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPGHIEIDFI 61
DB 1 MSAQKRLMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPGHIEIDFI 60
QY 62 RISGLRGKIGKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 121
DB 61 RISGLRGKIGKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 120
QY 122 VVLEHONGIAGLTNKLARIATKMQAEPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLEHONGIAGLTNKLAKIATKMQAEPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 241
DB 181 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 240
QY 242 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 301
DB 241 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 300
QY 302 AKIIEQPOLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAR 356
DB 301 AKIIEQPOLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAR 355

RESULT 4

AB0518
hypothetical protein STY0148 [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AB0518
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0518
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01285.1; PID:g16501413; GSPDB:GN00176
C:Genetics:
A:Gene: STY0148
C:Superfamily: murG protein

Query Match 89.2%; Score 1677; DB 2; Length 355;
Best Local Similarity 90.7%; Pred. No. 2.7e-116;
Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;
QY 2 MSGQKRLMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPGHIEIDFI 61
DB 1 MSAQKRLMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPGHIEIDFI 60
QY 62 RISGLRGKIGKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 121
DB 61 RISGLRGKIGKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 120

QY 122 VVLEHONGIAGLTNKLARIATKMQAEPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLEHONGIAGLTNKLAKIATKMQAEPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 241
DB 181 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 240
QY 242 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 301
DB 241 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 300
QY 302 AKIIEQPOLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAR 355
DB 301 AKIIEQPOLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAR 354

RESULT 5

AH0068
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosaminidase
C:Species: Versinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH0068
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Farrar, S.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 523-527, 2001
A:Title: Genome sequence of Versinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AH0068
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-356 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89411.1; PID:g15978647; GSPDB:GN00175
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 78.2%; Score 1471; DB 2; Length 356;
Best Local Similarity 78.3%; Pred. No. 4.2e-101;
Matches 278; Conservative 36; Mismatches 41; Indels 0; Gaps 0;

QY 2 MSGQKRLMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPGHIEIDFI 61
DB 1 MSGKTKRLMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPGHIEIDFI 60
QY 62 RISGLRGKIGKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 121
DB 61 KISGLRGKIGKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 120
QY 122 VVLEHONGIAGLTNKLARIATKMQAEPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLEHONGIAGLTNKLARIATKMQAEPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 241
DB 181 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 240
QY 242 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 301
DB 241 VEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 300
QY 302 AKIIEQPOLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAR 356
DB 301 AKIIEQPOLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAR 355

RESULT 6

E82081
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosaminidase
C:Species: Vibrio cholerae

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C/Accession: E82081
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
C/haxdson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000

A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: E82081
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-354 #HEI>
A/Cross-references: GB:AE004310; GB:AE003852; NID:G9656963; PIDN:AAF95544.1; GSPDB:GN001
A/A/Experimental source: serogroup O1; strain N16961; biotype El Tor
C/Genetics:
A/Gene: VC2401
A/Map position: 1
C/Superfamily: murG protein

Query Match 61.4%; Score 1154; DB 2; Length 354;
Best Local Similarity 64.2%; Pred.No.1e-77;
Matches 224; Conservative 48; Mismatches 77; Indels 0; Gaps 0;

Qy 2 MSGGQKILMVMWAGTGTGHVPFGLAVAHLMACQGWVRWLTGADRMEDADLPKHGIEIDFI 61
Db 1 MNKNKKLMVMWAGTGTGHVPFGLAVAKQLQQQWQIRWLTGADRMEDADLPKHGIEIDFI 60

Qy 62 RISGLRGKIGKALIAAPLRIFNAWRQARAIMKAYKPDVILGMGGYVSGPGGLAANSLGIP 121
Db 61 QVKGRLRGQGLMRLILKAPFQIVNAILQARRHLITYPQDAVLGMGGYVSGPGGIAAANLGLIP 120

Qy 122 VVLHEQNGIAGLTNKLARIATKVMQAEPCAFPMNAEVVGNPNVRTDVLALPLPOORLAGRE 181
Db 121 VVLHEQNAVAGLTNQLAKTARRVQFAFGAFADASVVGPNPVQDVQLAAPQRATEN 180

Qy 182 GPVRVLVVGSGQGARILINQTMPOVAAKGLDSVLIWHQSGKSGQSQSVQAEVAGQPHKV 241
Db 181 GATRILVWGSQGARILINQTLPAVMAALGSGYEIRHQAGKNSQQDVAAEYAAAGVESAQV 240

Qy 242 TERTDDMAAAYAVADVVCVRSGALTVSEIAAAGLPALFVPFHQKDRQOQVYNALPLEKAGA 301
Db 241 TEFIDDDVADAYAWADLLICRSGALTVSEVSAAGVGALFIFPMHKDRQOALNADHLVACGA 300

Qy 302 AKIIEQPQLSVDAVANTLAGNSRETLTMAERARAAASIPATERVANEV 350
Db 301 AKMIEQPELSVEKLTQWVRELDRAQLLSMAQKQAQAKLDADKVKVAQAI 349

RESULT 7

D64185

UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-ace

C Species: *Haemophilus influenzae*

C Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_charge 26-Aug-1999

C Accession: D64185

R: Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.A.;

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A: Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A: Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.

A: Reference number: A64000; MUID: 95350630; PMID: 7542800

A: Accession: D64185

A: Status: nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-351 <TIGR>

A: Cross-references: GS:U32793; GB:I42023; NID: g1574683; PIDN: AAC22793.1; PID: g1574693; T

C: Genetics:

C: Gene: murG

C: Function:

A: Description: involved in murin or cell envelope biosynthesis

A: Pathway: peptidoglycan biosynthesis

C: Superfamily: murG protein

C: Keywords: cell division; cell wall; glycosyltransferase; hexosyltransferase; peptidogl

```

Query Match          56.6%; Score 1064.5; DB 2; Length 351;
Best Local Similarity 61.5%; Pred. No. 4e-71;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;

QY      5  QGKRLWMAGGTGGHVFPGLAVAHHLMAQGQVQWRLGTADREADLPKPHGIEIDFFIRIS 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2  KKKLLWMAGGTGGHVFPFAIAVAQTLOQEWDIQWLGTKDRMEAQLVPKYGIPRFFIQIS 61

QY      65  GURGGIKALTAAPRIIFNANQRAINAKYKPDVVLGMGGYVSGPGGLAWSIGIPVVL 124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62  GURGGIKALKLNAPPFAIFRAVLQAKKIIOEBKPDVVLGMGGYVSGPGVAALKCGVFIL 121

QY      125  HEQNGIAGLTNKKLARIATKVMQREPGFAFNAEYVVGPNVRTDVLALPLPQORLAGRGPV 184
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122  HEONAIAGLTNKKLKIAITCVLOAPPTAFPAEYVVGPNVREDLPFEMPENPDIRFSDREKL 181

QY      185  RVLVVGGSGGARILNQTWPOVAAKLGDSVIIWHQSGKGSQSQVEQAVAEAGQPQHKVTEF 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      192  RVLVVGGSGGARVLNHTPLPKVVAQLADKLEFRHVGKGAVEVSQLYGE-NLEQVKITEF 240

QY      245  IDDMAAYAWADVVVCRSGALTVSEIAAAGLPALVPFPQHKDROQYNNALPLEKAGAAKI 304
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      241  IDNMAEYAWADVVICRSGALTVCEIAAAGAAAFVFPQHKDROQYNNALKLVSDVGAKEI 300

QY      305  IEQPQLSYDVAANTLAGHSRETLTMAERARAAASIPDATERVANEVSR 352
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      301  IEQADLTPEILVNYLKNLTRENLQMLAKKATMSMPNAAQRAEVIKQ 348

RESULT 8
D84955
hypothetical protein murG murG [imported] - Buchnera sp. (strain APS)
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: D84955
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. Af
A:Reference number: A84930; MUID:20445173; PMID:10993077
A:Accession: D84955
A:Status: preliminary
A:Accession: D84955
A:Molecule type: DNA
A:Residues: 1-354 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: murG; BU216
C:Superfamily: murG protein

```

Query Match	49.9%	Score 939	DB 2	Length 354
Best Local Similarity	48.4%	Pred. No. 7.3e-62		
Matches 169	Conservative 80	Mismatches 100	Indels 0	Gaps 0
Qy	7	KELVMWAGTGCHVFPGLVAVAHLMAQGWVRVLGTADRMEDLVKPHGIEIDFIRISGL	66	
Db	4	KIIIMAGSGGHVFPGLTIARVLIKGLVNMVIGTKNSIESRIIPKYGIKTHYISIKGL	63	
Qy	67	RKGGTKALIAPLRIFINAWROQRAIMKAYKPDVVLGMGGVYSGPGGLAANSIGIPVULHE	126	
Db	64	RNTSLKNIILIPVILRAYAVAKKIIETWSPDIVLGMGGVYSGPGGVAWNCNIIPLLLHE	123	
Qy	127	QNGIAGLTNKMARIATKWMQAEFGAFNAEVVGNPVRTDVLALPLPQORLACRGGFVRV	186	
Db	124	QNKIAGITNKLRSITKNMQASPGLVLRNAEVVGNPVCOSIIKVPNPINRFKNRTGLLRV	183	
Qy	187	LVVGSQSGARTILNOMFPQVAAKLGDSVIIHQSGKSGQSQSVQAYAAEGOPHKVTEFID	246	
Db	184	LVVGSQSSILNRIILPEVSPFLLEKIIIFHQGTGNYELEKTKYKYNKLRLNQNLITSPFK	243	
Qy	247	DMAAYAWADVVCRSGLTVTSEITAAAGLPALVPFQHKDRQOYWNALPIEKAGAAKIE	306	
Db	244	NTASAYEAWADJIIICRSGLTVTSEISIVGLGAIPIFPYHKDKOQHRNAEDLLIGAAKIID	303	

Best Local Similarity 46.2%, Pred. No. 1.4e-49;
Matches 160; Conservative 71; Mismatches 110; Indels 5; Gaps 4;

QY 6 GKRLMWAGTGGHVFGLAVAHLMQAQGVWLTGADMEADLVPKHGIEIDFIRISG 65
DB 3 GKTFLMAGTGGHVFGLAVADSLRARGHHVILWLGSKSMEEIRIVPQDILLETIAIG 62

RESULT 13
DB7023
hypothetical protein murG [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: DB7023
R:Colo, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21125732; PMID:11234002
A:Accession: DB7023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <STO>
A:Cross-references: GB:AL450380; NID:gl3092984; PIDN:CAC31295.1; GSPDB:GN00147
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 27.1%; Score 510.5; DB 2; Length 407;
Best Local Similarity 37.4%; Pred. No. 3.6e-30;
Matches 137; Conservative 62; Mismatches 138; Indels 29; Gaps 9;

QY 9 LMVAGGTGGHVFGLAVAHLMQAQGVWLTGADMEADLVPKHGIEIDFIRISGL 66
DB 31 VVLGGTGGHVFGLAVAHLMQAQGVWLTGADMEADLVPKHGIEIDFIRISGL 90

QY 67 RGKGIKALIAAPLRIFNARQARIMKAYKPDVVLGMGGYVSGPGLAAMSL----GIP 121
DB 91 PKLTGDLARLPLRVRAVRETRAVFEVVEAHVVGFGYVALPAYLAARGIPVRRIP 150

QY 122 VVLHONGIAGLTNKLWARIATKVMQAPGA-FPNAEVWGNVPTDVLALPLPQORLAGR 180
DB 151 VVVEANARAGIANRVGRTAERVLAVGSGLRGAEEVGVPIHATITTLNRPALRADAR 210

QY 181 E-----GPVRVLWV-GSGGARIINQTMPOVAAKLGDS-VIIHQSGKSGQSQSQVEQAYAE 233
DB 211 KHFGTDDARVLLVFGSGGAVSLNRVAGAEADLAAGVAVLH--AYGLKNTLELRTP 268

QY 234 AGQPOHKVTEFIDMAAAYAWADVVCRSGLATVSEIAAAGLPALFVPFHQKDRQYVNA 293
DB 269 YGEPYVAVPYLDRLMDLAYAAADLVICSGAMTVAEVSAGVLPAYVFPFPGNGEQLNA 328

QY 294 LPLEKAGAAKTIEQPLQSDAVANTLWAGSRETLLTMAERARASIPDATERV-----AN 348
DB 329 LDVWAGGLVVDADLTPLGLV-----RQVVLSDPQLAAMTAAARVGHDAH 381

QY 349 EVSRVA 354
DB 382 HVAKVA 387

RESULT 14
E70579
probable murG protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70579
R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Church, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

Best Local Similarity 46.2%, Pred. No. 1.4e-49;
Matches 160; Conservative 71; Mismatches 110; Indels 5; Gaps 4;

QY 6 GKRLMWAGTGGHVFGLAVAHLMQAQGVWLTGADMEADLVPKHGIEIDFIRISG 65
DB 3 GKTFLMAGTGGHVFGLAVADSLRARGHHVILWLGSKSMEEIRIVPQDILLETIAIG 62

RESULT 13
DB7023
hypothetical protein murG [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: DB7023
R:Colo, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21125732; PMID:11234002
A:Accession: DB7023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <STO>
A:Cross-references: GB:AL450380; NID:gl3092984; PIDN:CAC31295.1; GSPDB:GN00147
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 27.1%; Score 510.5; DB 2; Length 407;
Best Local Similarity 37.4%; Pred. No. 3.6e-30;
Matches 137; Conservative 62; Mismatches 138; Indels 29; Gaps 9;

QY 9 LMVAGGTGGHVFGLAVAHLMQAQGVWLTGADMEADLVPKHGIEIDFIRISGL 66
DB 31 VVLGGTGGHVFGLAVAHLMQAQGVWLTGADMEADLVPKHGIEIDFIRISGL 90

QY 67 RGKGIKALIAAPLRIFNARQARIMKAYKPDVVLGMGGYVSGPGLAAMSL----GIP 121
DB 91 PKLTGDLARLPLRVRAVRETRAVFEVVEAHVVGFGYVALPAYLAARGIPVRRIP 150

QY 122 VVLHONGIAGLTNKLWARIATKVMQAPGA-FPNAEVWGNVPTDVLALPLPQORLAGR 180
DB 151 VVVEANARAGIANRVGRTAERVLAVGSGLRGAEEVGVPIHATITTLNRPALRADAR 210

QY 181 E-----GPVRVLWV-GSGGARIINQTMPOVAAKLGDS-VIIHQSGKSGQSQSQVEQAYAE 233
DB 211 KHFGTDDARVLLVFGSGGAVSLNRVAGAEADLAAGVAVLH--AYGLKNTLELRTP 268

QY 234 AGQPOHKVTEFIDMAAAYAWADVVCRSGLATVSEIAAAGLPALFVPFHQKDRQYVNA 293
DB 269 YGEPYVAVPYLDRLMDLAYAAADLVICSGAMTVAEVSAGVLPAYVFPFPGNGEQLNA 328

QY 294 LPLEKAGAAKTIEQPLQSDAVANTLWAGSRETLLTMAERARASIPDATERV-----AN 348
DB 329 LDVWAGGLVVDADLTPLGLV-----RQVVLSDPQLAAMTAAARVGHDAH 381

QY 349 EVSRVA 354
DB 382 HVAKVA 387

RESULT 14
E70579
probable murG protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70579
R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Church, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70579
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-410 <COL>
A:Cross-references: GB:Z93388; GB:AL123456; NID:93261759; PIDN:CAB08640.1; PID:G2104325
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 26.9%; Score 506; DB 2; Length 410;
Best Local Similarity 37.8%; Pred. No. 7.7e-30;
Matches 143; Conservative 55; Mismatches 32; Gaps 10;

QY 3 SGQKELMWAGTGGHVPGLVAHLMACQWQVR--WLGTDADMEADLVKKGIEIDF 60
DB 32 SADSUSVVLGGTAGHVEPANAVALDPRVIRITAGTLRGLETWLPQGYHLEL 91
QY 61 IRISLGRGKIGKALIAAPLIRFNWQRAIMKAYKPDVVLGMGVGSGPGLAWSL-- 118
DB 92 ITAVPMRPRFGDGLARLPSPVRVRAVEARDVLDVADVVVGGYVLPAYLAARGPL 151
QY 119 -----GIPVLEHQNGIAGLTKWLAIAATKVMQAE--GAPNAEVGNPRTDVLALP 171
DB 152 PRRRRRIPVTHEANARAGLANRGAHTADVLSAVDSGLRRRAEVGVVPRASIAALD 211
QY 172 LPQORLAGR-----EGPVRVLV--GSGQARTINOTMPOVAAKLGD--VIIHQSGKGSQ 224
DB 212 RAVLRAEARAHFGFPDDARVLLVFGSGQAVSLNRAVSGAAADLAAGVCVLH--AHGFPQ 269
QY 225 QVQAYAEAGQPOHKVTEFFIDDMAAYAWADVVCRSALTIVSEIAAGLPALFPVPHQ 284
DB 270 NVLELRRRAGGPPYVAVPYLDRMELAYAAADLVCRAGMTVAEVSAGVPAIYVPLPI 329
QY 285 KDRQQVNNALPLEKAGAKIIIEQPOLSDVDVANTLAGSRETLTMAERARAASIPDATE 344
DB 330 GNGEQELNALPVNAGGGMVADAALTPELVARQVAG-----LLT--DPAALAATAAA 382
QY 345 RV-----ANEVSRVAPAL 357
DB 383 RVGHRDAAGQVAAALAV 400

RESULT 15
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70579
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-410 <COL>
A:Cross-references: GB:Z93388; GB:AL123456; NID:93261759; PIDN:CAB08640.1; PID:G2104325
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 26.7%; Score 501.5; DB 2; Length 380;
Best Local Similarity 36.7%; Pred. No. 1.5e-29;
Matches 135; Conservative 55; Mismatches 157; Indels 21; Gaps 10;

QY 2 MSQGLMWAGTGGHVPGLVAHLMACQWQVRWLTG---ADMEADLVKKGIEI 58

DB 5 LANQGV--IVLAGGTGCHLFPFAEALAHFARAGWDVH--LATDARAQRVFGAFADH--V 59
QY 59 DFRISLRGKIGKALIAAPLIRFNWQRAIMKAYKPDVVLGMGVGSGPGLAWSL 118
DB 60 HVIRSATIAGNPNVALKTFWISQWGLNDRKULFRRLKPKLVVGGYPTLPPLPYASNM 119
QY 119 GIPVLEHQNGIAGLTKWLAIAATKVMQAE--GAPNAEVGNPRTDVLALPLP 173
DB 120 GIPTLIHQNAVGRANKGLAGRVKAIAGGFLPENSAGAAKTVITGNPVRSPVLAAT 179
QY 174 QORLAGREGPVRVLVVGSGQARTINOTMPOVAAKLGD--VIIHQSGKGSQSQVEQ 229
DB 180 PYPAGKDDRRLLVFGSGQAOFFSOALPAAVALLPEHERARLLITQARKDEASARQ 239
QY 230 AYAEAGQPOHKVTEFFIDDMAAYAWADVVCRSALTIVSEIAAGLPALFPVPHQK-DRQ 288
DB 240 AYEKLGVPA--DVAPFFNDMPARMADAHFVIARSASTVSEITVIGRPAMLVPPPHALDHD 298
QY 289 QYNNALPLEKAGAKIIIEQPOLSDVDVANTL--AGSRETLTMAERARAASIPDATE 346
DB 299 QAANAAALAAAGAEVVRQADLSQRLAEMLOSAMNELEQQAQAAKASVGVKPDARLL 358
QY 347 ANEVSRVA 354
DB 359 ADLAEATA 366

RESULT 16
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70579
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-361 <STO>
A:Cross-references: GB:AE005673; NID:G13424116; PIDN:AAK24522.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2551
C:Superfamily: murG protein

Query Match 25.7%; Score 484; DB 2; Length 361;
Best Local Similarity 37.2%; Pred. No. 2.7e-28;
Matches 139; Conservative 55; Mismatches 140; Indels 40; Gaps 11;
QY 7 KRLMWAGTGGHVPGLVAHLMACQWQVRWLTGADR--MEADLVKKGIEIDFIRIS 64
DB 3 KLAVVAGGTGGHMFPAQALAEALARGWRV-VLATDGRALYADKFPAE-----ERIALS 57
QY 65 GLRGKIGKAL--IAAPLIRFNWQRAIMKAYKPDVVLGMGVGSGPGLAWSLGPV 122
DB 58 AATAKSNDFLGMKAGFVVLQGVWEARAFKRLDPAVVVGGYPALPALLGALSQGRPT 117
QY 123 VLHEQNGIAGLTKWLAIAATKVMQAE--GAPNAEVGNPRTDVLAL-PLPQ 174
DB 118 VIHEQNAVLRVNRFLAPRVNEVACAPFLEKATPAKCAHVYGNPVRFPFVRALEFDP- 176
QY 175 QRLAGREGPVRVLVVGSGQARTINOTMPOVAAKLGDVIIHQSGKGSQSQVEQYAE 234
DB 177 --YLAPEVQLRVLTGSGQARLLSEIPEAVAKLPEM-----RGLKVFQQAAR 227
QY 235 GQPOHK-----VTEPIDMAAYAWADVVCRSALTIVSEIAAGLPALFPVPHQ- 283
DB 228 MEQARKVYRNAMVECEVAPFDRMDVAGYLRQSHLIVGRSGASTCTELAVAGRPISILPKI 287
QY 284 HKDRQQVNNALPLEKAGAKIIIEQPOLSDVDVANTLAGM--SRETLTMAERARAASIPD 341

Db 288 AADHDFRNFARLEBAGAAVCLELTDVDMAAALKALLSKFERLEKMAAGARSAKEN 347
Qy 342 ATERVANEVSERVAR 355
Db 348 AABELADLVETKAR 361

RESULT 17
E83970
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)pyrophos murG [imported] - Bacill
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: E83970
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: E83970
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-363 <STO>
A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06284.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 25.0%; Score 470; DB 2; Length 363;
Best Local Similarity 32.5%; Pred. No. 3e-27;
Matches 123; Conservative 78; Mismatches 131; Indels 46; Gaps 12;

Qy 8 RLMVAGGTGGHVPGLAVAHHLMA--QGVQVRLWGTADRMEDLVKPKHGIEDFIRISG 65
Db 2 KIVVSGGTGGHVPGLAVAHHLMA--QGVQVRLWGTADRMEDLVKPKHGIEDFIRISG 61

Qy 66 LRKGKIKALIAAPLIRFNARQARIMKAYKPDVVLGGVYSGPGGLAAMSIGIPVVLH 125
Db 62 FQRLSMENKTVVRFLRGTKRAKALLNEFKPDVIGTGGVGGVPPVYAAKLIKPTVTH 121

Qy 126 EONGIAGTNNKLA---RIATKVMQAEPCGFNAEVW--GNPVRTDVLALPLPQORLAG 179
Db 122 EQNSVPLGTNNKLSRYDRIAICFKEAE--AFFPKNVVFTGNPRAEVS-----GN 172

Qy 180 REGPVR-----VLVVGSGQARILNQT-----MPQVAALGDSVLIWHQSGKSGQ 225
Db 173 REEGLRSLGKPKKTKTVLIVGSGRGARINDAFWSILSDVKAKPYQFVYV---TGTVHYE 229

Qy 226 SVEQAYAEAGOPQHKVTE--FIDDMAAYANADVVCESGALTYSETAAGLPAFLVFPFH 284
Db 230 RVQEMKSGIQPENVIQPFTHNPDPVLSAVDLLIVARAGATTIAETALGLPSILPSPY 289

Qy 285 -KDRQYWNALPLEKAGAATIEQPOLSDVAVANTLAGWSRETLT-----MAERARAA 337
Db 290 VTNNHQEKNAALSKDAAILKESLITGRLLEDI----DDIWTPTGRDLDMKQAKAL 345

Qy 338 SIPDATERVANEVSERVAR 355
Db 346 GVPTAAEKHLMLREVAK 363

RESULT 18
T34954
probable UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecapre
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T34954
R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21583
A:Accession: T34954
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-364 <SAU>
A:Cross-references: EMBL:AL109663; PIDN:CABS1993.1; GSPDB:GN00070; SCOEDB:SC4A10.17C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: murG; SCOEDB:SC4A10.17C
C:Superfamily: murG protein

Query Match 24.0%; Score 450.5; DB 2; Length 364;
Best Local Similarity 34.6%; Pred. No. 8.2e-26;
Matches 120; Conservative 58; Mismatches 142; Indels 27; Gaps 7;

Qy 9 LMMVAGGTGGHVPGLAVAHHLMAQGHQV--RWLGTADRMEDLVKPKHGIEDFIRISGL 66
Db 3 VVLAGGTGGTAGHIEPALADALRRQDPTVGTITALTGTERGLETRLVPERGYELALIPAVPL 62

Qy 67 RKGKIKALIAAPLIRFNARQARIMKAYKPDVVLGGVYSGPGGLAAMSIGIPVVLHE 126
Db 63 PRKPTPELITVPGRLRGITKATEQILERTKADAVAGGVYVALPAYLAARLGLVPIVHHE 122

Qy 127 QNGIAGTNNKLAIRATKVMQAEPCG--FPNAEVGNVVRTDVLALPLPQORLAGR----- 180
Db 123 ANARPLANKIGSRVAAQVAVSTPDSKLRNSRYIGIFLRSIATLDRAAARPERAMFGL 182

Qy 181 -EGPVRLVVGSGQARILNQTMPVAAKLGDSVI--IWHQSGKSGQSVQAYAEAGOPQ 238
Db 183 DPNLPTLLVVGSGQARILNEVIQQVAPWLQQAQIGIQLH--AVGPKNELPQVHQMGPMP 240

Qy 239 HKVTEFIDDMAAAYAWADVVCESGALTYSBIAAGLPAFLVFPFHKDRQYWNALPLEK 298
Db 241 YIPVSYLDRMDLAVADMLCRAGMTVAELSAVGIPAAVYVPLPIGNGBQLNAQPVVK 300

Qy 299 AGAKIKIQPOLSDVAVANTLAGWSRETLT-----TMAERARAA 338
Db 301 AGGGLLVDDAELTPE-----WLQNVLPVLPADPHRLVEMSRAAA 339

RESULT 19
AG2833
hypothetical protein murG [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AG2833
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AG2833
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <KUR>
A:Cross-references: GB:AB008688; PIDN:AAL43085.1; PID:g17740555; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: murG
A:Map position: circular chromosome
C:Superfamily: murG protein

Query Match 23.5%; Score 442; DB 2; Length 378;
Best Local Similarity 33.4%; Pred. No. 3.7e-25;
Matches 126; Conservative 59; Mismatches 140; Indels 52; Gaps 13;

Qy 9 LMMVAGGTGGHVPGLAVAHHLMAQGHQV-----RWLGTADRMEDLVKPKHGIED 59
Db 6 VLLAAGGTGGHVPGLAALAHKATKARGYQVHLVTDPSAERYAKGFPADSIHVVPSATI--- 62

Qy 60 FIRISLGRKGKIKALIAAPLIRFNAAW---ROARATMKAYKPDVVLGGVYSGPGGLAAW 116
Db 63 -----GSKNPFISVRSILKWLWGLRTARLVTKLKPVAVGVFGYVTPVPLLAST 112

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 07:01:04 ; Search time 17 Seconds
(without alignments)
1114.914 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 MMSQGRKRLVMAGGTGGHV.....RVANEVSRVARALEHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1798	95.6	354	1	MURG_ECOLI
2	1793	95.3	355	1	MURG_SHIFL
3	1788	95.1	354	1	MURG_ECOS7
4	1786	94.9	354	1	MURG_ECOS6
5	1682	89.4	354	1	MURG_SALTY
6	1672	88.9	354	1	MURG_SALTI
7	1471	78.2	356	1	MURG_YERPE
8	1431	76.1	355	1	MURG_PHOLL
9	1175	62.5	355	1	MURG_VIBPA
10	1166	62.0	355	1	MURG_VIBVU
11	1163	61.8	355	1	MURG_VIBVY
12	1154	61.4	354	1	MURG_VIBCH
13	1123	59.7	354	1	MURG_PASWU
14	1064.5	56.6	351	1	MURG_HAEIN
15	1021	54.3	355	1	MURG_HAEUD
16	945	50.2	354	1	MURG_BUCAP
17	939	49.9	354	1	MURG_BUCAI
18	900	47.8	360	1	MURG_CANBF
19	887	47.2	362	1	MURG_SHRON
20	869.5	46.2	365	1	MURG_RALSO
21	863.5	45.9	365	1	MURG_SHEVI
22	801	42.6	353	1	MURG_WUCBP
23	798	42.4	359	1	MURG_WIGBR
24	798	42.4	364	1	MURG_XYLFT
25	794	42.2	357	1	MURG_BORPA
26	793	42.2	357	1	MURG_BORBR
27	791.5	42.1	426	1	MURG_XANAC
28	791	42.1	357	1	MURG_BORPE
29	787	41.8	358	1	MURG_CORXB
30	783.5	41.7	427	1	MURG_XANCP
31	780	41.5	367	1	MURG_XYLFA
32	779.5	41.4	355	1	MURG_NEIMB
33	772.5	41.1	355	1	MURG_NEIMA

RESULT 1

34	764	40.6	1	MURG_NITEU
35	719.5	38.3	1	MURG_PSEAE
36	717	38.1	1	MURG_PSESM
37	683	36.3	1	MURG_PSEPK
38	510.5	27.1	1	MURG_MICLE
39	506	26.9	1	MURG_MYCTO
40	506	26.9	1	MURG_MYCTO
41	501.5	26.7	1	MURG_BRUME
42	496.5	26.4	1	MURG_BRUSU
43	487	25.9	1	MURG_CORGL
44	484	25.7	1	MURG_CAUCR
45	480.5	25.5	1	MURG_COREF
46	480	25.5	1	MURG_RHIME
47	476	25.3	1	MURG_BACTN
48	470	25.0	1	MURG_BACHD
49	470	25.0	1	MURG_RHILO
50	469	24.9	1	MURG_PORGI
51	462	24.6	1	MURG_BACCR
52	461.5	24.5	1	MURG_STRAW
53	460	24.5	1	MURG_BACAA
54	451	24.0	1	MURG_SYNEL
55	450.5	24.0	1	MURG_STRCO
56	447.5	23.8	1	MURG_STRCU
57	442	23.5	1	MURG_AGRTS
58	440	23.4	1	MURG_GLOVI
59	438	23.3	1	MURG_THETN
60	435	23.1	1	MURG_LACLA
61	433	23.0	1	MURG_ZYMMO
62	431.5	22.9	1	MURG_BIFLO
63	425	22.6	1	MURG_ANASP
64	424	22.5	1	MURG_LACPL
65	422	22.4	1	MURG_BACSU
66	422	22.4	1	MURG_LISMO
67	419	22.3	1	MURG_ENTHR
68	418	22.2	1	MURG_LISIN
69	411.5	21.9	1	MURG_ENTFA
70	407	21.6	1	MURG_CHLTE
71	404	21.5	1	MURG_BRAJA
72	398.5	21.2	1	MURG_SRYA3
73	340	18.1	1	MURG_RICCN
74	335	17.8	1	MURG_OCEIH
75	334.5	17.8	1	MURG_AQAEH
76	332.5	17.7	1	MURG_CLOTE
77	328.5	17.5	1	MURG_CLOAB
78	321.5	17.1	1	MURG_THEMA
79	319.5	17.0	1	MURG_RICPR
80	315.5	16.8	1	MURG_TREPA
81	313	16.6	1	MURG_CHLTR
82	306.5	16.3	1	MURG_BACCR
83	304.5	16.2	1	MURG_BACAA
84	304.5	16.2	1	MURG_CLOPE
85	303	16.1	1	MURG_CHLMU
86	299	15.9	1	MURG_CHLCV
87	284.5	15.1	1	MURG_CHLPN
88	277	14.7	1	MURG_HELPY
89	275	14.6	1	MURG_HELPY
90	273	14.5	1	MURG_STAAM
91	273	14.5	1	MURG_STAAM
92	272.5	14.5	1	MURG_STRRG
93	269.5	14.3	1	MURG_STRPN
94	263	14.0	1	MURG_CAMJE
95	261	13.9	1	MURG_STAEF
96	259.5	13.8	1	MURG_BORBU
97	250	13.3	1	MURG_STRPY
98	248	13.2	1	MURG_STRPY
99	213.5	11.4	1	MURG_STRMU
100	150	8.0	1	Y453_METAC

ALIGNMENTS

MURG_ECOLI STANDARD; PRT; 354 AA.

AC P1743; 1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 30, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)

DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase

DE (SC 2.4.1.227) (Undecaprenyl-PP-MurNAC-pentapeptide-UDPGlcNAC GlcNAC transferase).

DE MURG OR B0090.

GN Escherichia coli.

OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=90326550; PubMed=2197603;

RA Ikeda M., Wachi M., Jung H.K., Ishino F., Matsuhashi M.;

RT "Nucleotide sequence involving murg and murC in the mra gene cluster region of Escherichia coli.";

RL Nucleic Acids Res. 18:4014-4014(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=92334977; PubMed=1630901;

RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,

RT "Systematic sequencing of the Escherichia coli genome: analysis of the 0-2.4 min region.";

RL Nucleic Acids Res. 20:3305-3308(1992).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [4]

RP SEQUENCE OF 8-354 FROM N.A.

RC STRAIN=K12;

RX MEDLINE=90251463; PubMed=2187180;

RA Mengin-Lecreulx D., Texier L., van Heijenoort J.;

RT "Nucleotide sequence of the cell-envelope murg gene of Escherichia coli.";

RL Nucleic Acids Res. 18:2810-2810(1990).

RN [5]

RP SEQUENCE OF 1-17, AND CHARACTERIZATION.

RX MEDLINE=91310568; PubMed=1649817;

RA Mengin-Lecreulx D., Texier L., Rousseau M., van Heijenoort J.;

RT "The murg gene of Escherichia coli codes for the UDP-N-acetylglucosamine: N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis.";

RL J. Bacteriol. 173:4625-4636(1991).

RN [6]

RP SUBCELLULAR LOCATION.

RX MEDLINE=93194811; PubMed=8449890;

RA Bupp K., van Heijenoort J.;

RT "The final step of peptidoglycan subunit assembly in Escherichia coli occurs in the cytoplasm.";

RL J. Bacteriol. 175:1841-1843(1993).

RN [7]

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RX MEDLINE=20348887; PubMed=10892798;

RA Ha S., Walker D., Shi Y., Walker S.;

RT "The 1.9 A crystal structure of Escherichia coli Murg, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis.";

Protein Sci. 9:1045-1052(2000).

!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide)GlcNAC (lipid intermediate II).

!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(Oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP + GlcNAC-(1->4)-Mur2Ac(Oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.

!- PATHWAY: Peptidoglycan biosynthesis; last step.

!- SUBCELLULAR LOCATION: Inner membrane-associated.

!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X52644; CAA36867.1; -

DR EMBL; X52540; CAA36776.1; -

DR EMBL; X55034; CAA38867.1; -

DR EMBL; D10483; BAB96658.1; -

DR EMBL; AE000118; AAC73201.1; -

DR PIR; J00544; BVCEMG.

DR HAMAP; MF_00033; -; 1.

DR EcoGene; EG10623; murg.

DR InterPro; IPR007235; Glyco_tran_28_C.

DR InterPro; IPR004276; Glyco_trans_28.

DR InterPro; IPR006009; MurG_1.

DR Pfam; PF04101; Glyco_tran_28_C; 1.

DR Pfam; PF03033; Glyco_transf_28; 1.

DR TIGRFAMs; TIGR01133; murg; 1.

KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;

KW Inner membrane; Peptidoglycan synthesis; Complete proteome;

KW 3D-structure.

INIT_MET 0 0

STRAND 6 10

HELIX 15 29

TURN 30 32

TURN 34 39

TURN 41 42

TURN 44 44

TURN 45 48

TURN 49 52

TURN 53 53

TURN 55 58

TURN 65 66

TURN 69 73

TURN 74 74

TURN 76 93

TURN 97 100

TURN 104 105

TURN 106 115

TURN 116 117

TURN 120 124

TURN 131 136

TURN 137 139

TURN 142 145

TURN 148 149

TURN 155 156

TURN 163 166

TURN 167 167

TURN 171 175

TURN 176 177

TURN 182 187

TURN 189 191

TURN 194 207

TURN 208 210


```

FT STRAND 211 216
FT TURN 219 220
FT HELIX 231 232
FT TURN 232 233
FT TURN 233 234
FT TURN 235 236
FT STRAND 238 240
FT HELIX 246 252
FT STRAND 255 258
FT HELIX 262 271
FT TURN 272 277
FT STRAND 275 277
FT TURN 283 284
FT HELIX 286 296
FT TURN 297 298
FT STRAND 300 302
FT HELIX 305 307
FT TURN 310 318
FT TURN 319 319
FT HELIX 322 334
FT TURN 335 335
FT TURN 338 339
FT HELIX 340 352
FT TURN 353 354
SQ SEQUENCE 354 AA; 37683 MW; 0E3FAD945D769CID CRC64;

Query Match
Best Local Similarity 95.6%; Score 1798; DB 1; Length 354;
Matches 351; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 SQGQKRLVMAGGTGGHVPGLAVAHLMMAQGVQVRLGTADRMADLVPKHGIEDIFR 62
DB 1 SQGQKRLVMAGGTGGHVPGLAVAHLMMAQGVQVRLGTADRMADLVPKHGIEDIFR 60

QY 63 ISGLRGKIKALIAAPRIFNWQOARIMKAYKPDVVLGNGGVSGPGGLAANSLGIPV 122
DB 61 ISGLRGKIKALIAAPRIFNWQOARIMKAYKPDVVLGNGGVSGPGGLAANSLGIPV 120

QY 123 VLHEQNGIAGLTNKLWARIATKVMQABPGFAPNAEVVGNFVRTDVLALPLPQORLAGREG 182
DB 121 VLHEQNGIAGLTNKLWAKIATKVMQAFPGFAPNAEVVGNFVRTDVLALPLPQORLAGREG 180

QY 183 PVRVLVVGSGQARILNCTMPQVAAKLGDSVTIWHQSGKSGQSVQVQYAEAGQPOHKVT 242
DB 181 PVRVLVVGSGQARILNCTMPQVAAKLGDSVTIWHQSGKSGQSVQVQYAEAGQPOHKVT 240

QY 243 EFIDDMAAYAWADVVCRSALTIVSEIAAGLPALFVPFQHKDRQOYWNALPLEKAGAA 302
DB 241 EFIDDMAAYAWADVVCRSALTIVSEIAAGLPALFVPFQHKDRQOYWNALPLEKAGAA 300

QY 303 KIIQPQLSVDVANTLAGSRETLLTWABRARAASIPDATERVANEVSVARA 356
DB 301 KIIQPQLSVDVANTLAGSRETLLTWABRARAASIPDATERVANEVSVARA 354

RESULT 2
MURG_SHIFL STANDARD; PRT; 355 AA.
AC Q8JMN4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.22) (Undecaprenyl-P-P-MurNac-pentapeptide-UDP-GlcNAc
DE transferase).
GN MURG OR SFO087 OR S0089.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;

```

```

RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding X., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB015045; AAN1752.1; -
DR EMBL; AB016978; AAP15633.1; -
DR HAMAP; MF_00033; -; 1.
DR InterPro; IPR004276; Glyco_tran_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR006009; MurG_tran_28.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRFAMs; TIGR01133; murG; 1.
DR Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 355 AA; 37812 MW; 73407776C2B1504C CRC64;

Query Match
Best Local Similarity 95.3%; Score 1793; DB 1; Length 355;
Matches 350; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 MSGQKRLVMAGGTGGHVPGLAVAHLMMAQGVQVRLGTADRMADLVPKHGIEDIFR 61
DB 1 MSGQKRLVMAGGTGGHVPGLAVAHLMMAQGVQVRLGTADRMADLVPKHGIEDIFR 60

QY 62 RISGLRGKIKALIAAPRIFNWQOARIMKAYKPDVVLGNGGVSGPGGLAANSLGIP 121
DB 61 RISGLRGKIKALIAAPRIFNWQOARIMKAYKPDVVLGNGGVSGPGGLAANSLGIP 120

QY 122 VLHEQNGIAGLTNKLWARIATKVMQABPGFAPNAEVVGNFVRTDVLALPLPQORLAGRE 181
DB 121 VLHEQNGIAGLTNKLWAKIATKVMQAFPGFAPNAEVVGNFVRTDVLALPLPQORLAGRE 180

QY 182 GPVRLVVGSGQARILNCTMPQVAAKLGDSVTIWHQSGKSGQSVQVQYAEAGQPOHKV 241
DB 181 GPVRLVVGSGQARILNCTMPQVAAKLGDSVTIWHQSGKSGQSVQVQYAEAGQPOHKV 240

```

```
QY 242 TEIDDMAAYAVADVVCVRSGLTYSSEIAAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 301
Db 241 TEIDDMAAYAVADVVCVRSGLTYSSEIAAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 300
QY 302 AKIIEQPQLSVDVANTLACWSRETLTMAERARAASIPDATERVANEVSRVARA 356
Db 301 AKIIEQPQLSVDVANTLACWSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 3
MURG EC057 STANDARD; PRT; 354 AA.
AC Q8X9Y8;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNAC-pentapeptide-UDPGlcNAC GlcNAC
DE transferase).
GN MURG OR 20100 OR EGS0094.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Hackett J., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 403:529-533 (2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAC
CC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-
CC (pentapeptide)GlcNAC (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAC-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/submit/
CC or send an email to license@ebi.ac.uk).
```

```
DR PIR; F90640; F90640.
DR HAMAP; MF_00033; --; 1.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR006009; MurG-.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR TIGRFAMs; TIGR01133; murG; 1.
KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
FT INIT MET 0 BY SIMILARITY.
SQ SEQUENCE 354 AA; 37669 MW; 8C4617F0EBA10DE9 CRC64;

Query Match 95.1%; Score 1788; DB 1; Length 354;
Best Local Similarity 98.6%; Pred. No. 9.7e-124;
Matches 349; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 SQGKRLVMVAGGTGGHVFPGLAVAHHLMAQGVQVWLGTADRMEDLVPKHGIEIDFIR 62
Db 1 SAQKRLVMVAGGTGGHVFPGLAVAHHLMAQGVQVWLGTADRMEDLVPKHGIEIDFIR 60
QY 63 ISGLRGKGIKALIAAPLRFINAWRQARIMKAYKPDVVVLGMGVVSGPGLAANSLGIPV 122
Db 61 ISGLRGKGIKALIAAPLRFINAWRQARIMKAYKPDVVVLGMGVVSGPGLAANSLGIPV 120
QY 123 VLHEQNGIAGLTNKLWARIATKWAQEPGAPFNAEVVGNPVRTDLALPLPQORLAGREG 182
Db 121 VLHEQNGIAGLTNKLWARIATKWAQEPGAPFNAEVVGNPVRTDLALPLPQORLAGREG 180
QY 183 PVRVLVVGSGQCARILNQTMPQVAAKLGDSVTIWHQSGKSGSQSQVQAYABAGQPQHKVT 242
Db 181 PVRVLVVGSGQCARILNQTMPQVAAKLGDSVTIWHQSGKSGSQSQVQAYABAGQPQHKVT 240
QY 243 EPIDMAAYAWADVVCVRSGLTYSSEIAAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 302
Db 241 EPIDMAAYAWADVVCVRSGLTYSSEIAAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 300
QY 303 KIIEQPQLSVDVANTLACWSRETLTMAERARAASIPDATERVANEVSRVARA 356
Db 301 KIIEQPQLSVDVANTLACWSRETLTMAERARAASIPDATERVANEVSRVARA 354

RESULT 4
MURG EC0L6 STANDARD; PRT; 354 AA.
ID MURG EC0L6
AC Q8FL64;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNAC-pentapeptide-UDPGlcNAC GlcNAC
DE transferase).
GN MURG OR C0108.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388334; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAC
CC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-
CC (pentapeptide)GlcNAC (lipid intermediate II) (By similarity).
CC
```

CC CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A016755; AAN78606.1; -
CC HAMAP; MF_00033; -; 1.
CC InterPro; IPR007235; Glyco_tran_28.C.
CC InterPro; IPR004276; Glyco_trans_28.
CC InterPro; IPR006009; MurG.
CC Pfam; PF04101; Glyco_tran_28.C; 1.
CC Pfam; PF03033; Glyco_transf_28; 1.
CC TIGRFAMs; TIGR01133; murG; 1.
CC Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
CC Inner membrane; Peptidoglycan synthesis; Complete proteome.
CC INIT MET 0 BY SIMILARITY.
CC FT
CC SQ SEQUENCE 354 AA; 37645 MW; 0C74C03FDD769C1D CRC64;

Query Match 94.9%; Score 1786; DB 1; Length 354;
Best Local Similarity 99.6%; Pred. No. 1.4e-123;
Matches 349; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SQGKRLMVMAGGTGGHVFPGGLAVAHLMAGQVQVRLGTADRMEDLVPKHGIEDIFR 62
Db 1 SQGKRLMVMAGGTGGHVFPGGLAVAHLMAGQVQVRLGTADRMEDLVPKHGIEDIFR 60

Qy 63 ISGLRGKIKALIAAPLIRIFNANQARAIMKAYKPDVVLGMGGYVSGPGLAASLGIPV 122
Db 61 ISGLRGKIKALIAAPLIRIFNANQARAIMKAYKPDVVLGMGGYVSGPGLAASLGIPV 120

Qy 123 VLHQNGIAGLTNKLARIATKWAEPGAPPAEAVVGNPVRTDVLALPLPQORLAGREG 182
Db 121 VLHQNGIAGLTNKLAKIATVWQAPPAEAVVGNPVRTDVLALPLPQORLAGREG 180

Qy 183 PVRVLVVGSGQARILNQTMPQVAAKLGDSVIIWHQSGKGSQSQVQAYAEAGQPQHKVT 242
Db 121 VLHQNGIAGLTNKLAKIATVWQAPPAEAVVGNPVRTDVLALPLPQORLAGREG 180

Qy 183 PVRVLVVGSGQARILNQTMPQVAAKLGDSVIIWHQSGKGSQSQVQAYAEAGQPQHKVT 242
Db 181 PVRVLVVGSGQARILNQTMPQVAAKLGDSVIIWHQSGKGSQSQVQAYAEAGQPQHKVT 240

Qy 243 EFIDDMAAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDQOQVWALPLEKAGAA 302
Db 241 EFIDDMAAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDQOQVWALPLEKAGAA 300

Qy 303 KIIEQPQLSDVAVANTLAGNSRETLTMAERARAASIPDATERVANEVSRVAA 356
Db 301 KIIEQPQLSDVAVANTLAGNSRETLTMAERARAASIPDATERVANEVSRVAA 354

RESULT 5
MURG SALTY STANDARD; PRT; 354 AA.
AC O82RU3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.22) (Undecaprenyl-pp-MurNAC-pentapeptide-UDP-GlcNAc
DE transferase).
GN MURG OR STM0128
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
RT LT2";
RL Nature 413:852-856 (2001).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-
CC (pentapeptide/GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A008699; AAL19092.1; -
CC StyGene; SG????; murG.
CC HAMAP; MF_00033; -; 1.
CC InterPro; IPR007235; Glyco_tran_28.C.
CC InterPro; IPR004276; Glyco_trans_28.
CC InterPro; IPR006009; MurG.
CC Pfam; PF04101; Glyco_tran_28.C; 1.
CC Pfam; PF03033; Glyco_transf_28; 1.
CC TIGRFAMs; TIGR01133; murG; 1.
CC Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
CC Inner membrane; Peptidoglycan synthesis; Complete proteome.
CC INIT MET 0 BY SIMILARITY.
CC FT
CC SQ SEQUENCE 354 AA; 37731 MW; 0E3850B6C6DCFD2E CRC64;

Query Match 89.4%; Score 1682; DB 1; Length 354;
Best Local Similarity 91.2%; Pred. No. 5.5e-116;
Matches 322; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

Qy 3 SQGKRLMVMAGGTGGHVFPGGLAVAHLMAGQVQVRLGTADRMEDLVPKHGIEDIFR 62
Db 1 SQGKRLMVMAGGTGGHVFPGGLAVAHLMAGQVQVRLGTADRMEDLVPKHGIEDIFR 60

Qy 63 ISGLRGKIKALIAAPLIRIFNANQARAIMKAYKPDVVLGMGGYVSGPGLAASLGIPV 122
Db 61 ISGLRGKIKALIAAPLIRIFNANQARAIMKAYKPDVVLGMGGYVSGPGLAASLGIPV 120

Qy 123 VLHQNGIAGLTNKLARIATKWAEPGAPPAEAVVGNPVRTDVLALPLPQORLAGREG 182
Db 121 VLHQNGIAGLTNKLAKIATVWQAPPAEAVVGNPVRTDVLALPLPQORLAGREG 180

Qy 183 PVRVLVVGSGQARILNQTMPQVAAKLGDSVIIWHQSGKGSQSQVQAYAEAGQPQHKVT 242
Db 181 PVRVLVVGSGQARILNQTMPQVAAKLGDSVIIWHQSGKGSQSQVQAYAEAGQPQHKVT 240

Qy 243 EFIDDMAAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDQOQVWALPLEKAGAA 302
Db 241 EFIDDMAAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDQOQVWALPLEKAGAA 300

QY 303 KITEPOLSDVAVANTLACGSRRETLTMAERAPASIPDATERVANEVSRRVAR 355
 Db 301 KIFEQCFTEVAVADTLACGSRRETLTMAERAPASIPDATERVANEVSRRVAR 353

RESULT 6

MURG_SALTI STANDARD; PRT; 354 AA.
 AC Q829G9; 2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
 DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
 DE (EC 2.4.1.227) [Undecaprenyl-PP-MurNAc-pentapeptide-UDP-GlcNAc GlcNAc transferase].
 GN MURG OR STY0148 OR T0132.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Holdren M.T.G., Sebahia M.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Connor P.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Rarrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsis K.,
 RA Krogh A., Larsen T.S., Leather S., Moulis S., O'Gaora P., Farry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 enterica serovar Typhi CT18";
 RT Nature 413:848-852(2001).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
 RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
 subunit on undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide (lipid
 intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAc-
 (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(cyl-L-Ala-
 GlcNAc-(1->4)-Mur2Ac(cyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
 diphosphoundecaprenol.
 CC -1- PATHWAY: Peptidoglycan biosynthesis; last step.
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
 subfamily.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AL627265; CADD1285.1; .
 EMBL; AE016834; AAG67864.1; .
 HMAP; MF 00033; .; 1.
 InterPro; IPR007235; Glyco_tran_28.C.
 InterPro; IPR004276; Glyco_tran_28.
 InterPro; IPR006009; MurG.
 Pfam; PF04101; Glyco_tran_28.C; 1.

DR Pfam; PF03033; Glyco transf_28; 1.
 DR TIGRfams; TIGR01133; murG; 1
 KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
 KW inner membrane; Peptidoglycan synthesis; Complete proteome.
 FT INIT MET 0 BY SIMILARITY.
 SQ SEQUENCE 354 AA; 37689 MW; 9BE582B5E69BBD72 CRC64;
 Query Match 88.9%; Score 1672; DB 1; Length 354;
 Best Local Similarity 90.7%; Pred. No. 3e-115;
 Matches 320; Conservative 19; Mismatches 14; Indels 0; Gaps 0;
 QY 3 SQGKRLVMAGTGGHVFPGCLAVAHLMACGQVWRMLGTADNRMEADLVPHKGIIDFIR 62
 Db 1 SQGPKRLVMAGTGGHVFPGCLAVAHLMACGQVWRMLGTSDRMEADLVPHKGIIDFIR 60
 QY 63 ISGLRGKIGKIALIAPLRIFNANRQARAIMKAYKPDVVLGMMGGVSGPGLAAMSLGIPV 122
 Db 61 ISGLRGKGVKALLAAPLRIFNANRQARAIMKRFKPDVVLGMMGGVSGPGLAAMSLGIPV 120
 QY 123 VLHEQNGIAGLTNKLWARIATKVMQAFPGAPNAEVVGNFVRDVLALPLPQRLAGREG 182
 Db 121 VLHEQNGIAGLTNQWLAKIATTVNQAFPGAPNAEVVGNFVRDVLALPLPQRLAGRDG 180
 QY 183 PVRVLVVGSSQCARILNQTPQVAAKLGDSVLIWHQSGKSQSQSVQAYAEAGOPQHKVT 242
 Db 181 PIRVLVVGSSQCARVLNQTPQVAARLGDTVTIWHQSGKAQLTVQAYAGAGOPQHKVT 240
 QY 243 EFIDDMAAAAYAWADVVCRCGALTVSEIAAAGLPALFPFQHKDRQQYWNALPLEKAGAA 302
 Db 241 EFIDGMAAAYAWADVVCRCGALTVSEIAAAGLPALFPFQHKDRQQYWNALPLENAGAA 300
 QY 303 KIIPQOLSVDVANTLAGWSRETLTMAERAPASIPDATERVANEVSRRVAR 355
 Db 301 KIFEQPOFTVEAVADTLACGSRRETLTMAERAPASIPDATERVANEVSRRVAR 353

RESULT 7

MURG_YERPE STANDARD; PRT; 356 AA.
 AC Q82IE9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
 DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
 DE (EC 2.4.1.227) [Undecaprenyl-PP-MurNAc-pentapeptide-UDP-GlcNAc GlcNAc transferase].
 DE MURG OR YPO0555 OR Y3626.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagsis K., Kariyeh A.V.,
 RA Leather S., Moulis S., Oyston P.C.F., Quail M.A., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
 RA "Genome sequence of Yersinia pestis, the causative agent of plague";
 RL Nature 413:523-527(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,

Query Match 62.0%; Score 1166; DB 1; Length 355;
Best Local Similarity 65.9%; Pred. No. 3.2e-78;
Matches 228; Conservative 47; Mismatches 71; Indels 0; Gaps 0;

QY 5 QGKRLMWAGGTGGHVPFGLAVAHHLMAQGMQVWLGTADRMADLVPKHGIEIDFIRIS 64
DB 3 KNKRLMWAGGTGGHVPFGLAVAKKLOQGWIRWLTADRMADLVPKHGIDIDIKVK 62

QY 65 GLRGKIKALIAAPLRIFNWQRAIMKAYKPDVVLGMGVGSGGGLAANSLGIPVVL 124
DB 63 GLRGQIKRLVLAPFQILNAIFQAKHAKRQWPDVVLGMGVGSGGGLAANSLGIPVVL 122

QY 125 HEQNGIAGLTNKLWARIATKVMQAPFGAPFNAEVVGNPVRTDLALPLPQORLAGREGPV 184
DB 123 HEQNAVAGLTNHLWAKIAKVFQAFPGAFKQAPVGNPVRDVALPDMQRMQDREGAV 182

QY 185 RVLVVGSGQARILNQTMPQVAAKLGDVLIWHQSGKSGQSQVEQAYAEAGQPHKYTEF 244
DB 183 RILVVGSGQARILNQTMPQVAAKLGDVLIWHQSGKSGQSQVEQAYAEAGQPHKYTEF 242

QY 245 IDMAAAYAWADVVCRSGALTVSEIAAGLPALFVFPQHKDQOQYVNALPLEKAGAARI 304
DB 243 IDVAAQYAWADLVCRSGALTVSEVSAAGVGAIFIPFMHQRQOALNADHLVACGAALM 302

QY 305 IEQPOLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEV 350
DB 303 IEQPOLTVDLKAGIQKLGRTLLSMALHARAAQNNADQVVADAI 348

RESULT 11
MURG_VIBVY STANDARD; PRT; 355 AA.

AC Q7NNVL;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNAc
DE transferase).
GN MURG OR VV0614.
OS Vibrio vulnificus (strain YJ016).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OC NCBI_TaxID=196600;
RN [1]
RP Chen C.Y., Wu K.M., Chang Y.C., Chang C.H., Tsai H.C., Liao T.L.,
RA Liu Y.M., Chen H.J., Shen A.B., Li J.C., Su T.L., Shao C.P., Lee C.T.,
RA Hor L.I., Tsai S.F.;
RT "Comparative genome analysis of Vibrio vulnificus, a marine
pathogen."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)/GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

CC EMBL; AP05332; BAC93378.1; ALT_INIT.
DR HAMAP; MF_00033; -; 1.
KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 355 AA; 38170 MW; 1CED565A38A9D120 CRC64;

Query Match 61.8%; Score 1163; DB 1; Length 355;
Best Local Similarity 65.9%; Pred. No. 5.3e-78;
Matches 228; Conservative 46; Mismatches 72; Indels 0; Gaps 0;

QY 5 QGKRLMWAGGTGGHVPFGLAVAHHLMAQGMQVWLGTADRMADLVPKHGIEIDFIRIS 64
DB 3 KNKRLMWAGGTGGHVPFGLAVAKKLOQGWIRWLTADRMADLVPKHGIDIDIKVK 62

QY 65 GLRGKIKALIAAPLRIFNWQRAIMKAYKPDVVLGMGVGSGGGLAANSLGIPVVL 124
DB 63 GLRGQIKRLVLAPFQILNAIFQAKHAKRQWPDVVLGMGVGSGGGLAANSLGIPVVL 122

QY 125 HEQNGIAGLTNKLWARIATKVMQAPFGAPFNAEVVGNPVRTDLALPLPQORLAGREGPV 184
DB 123 HEQNAVAGLTNHLWAKIAKVFQAFPGAFKQAPVGNPVRDVALPDMQRMQDREGAV 182

QY 185 RVLVVGSGQARILNQTMPQVAAKLGDVLIWHQSGKSGQSQVEQAYAEAGQPHKYTEF 244
DB 183 RILVVGSGQARILNQTMPQVAAKLGDVLIWHQSGKSGQSQVEQAYAEAGQPHKYTEF 242

QY 245 IDMAAAYAWADVVCRSGALTVSEIAAGLPALFVFPQHKDQOQYVNALPLEKAGAARI 304
DB 243 IDVAAQYAWADLVCRSGALTVSEVSAAGVGAIFIPFMHQRQOALNADHLVACGAALM 302

QY 305 IEQPOLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEV 350
DB 303 IEQPOLTVDLKAGIQKLGRTLLSMALHARAAQNNADQVVADAI 348

RESULT 12
MURG_VIBCH STANDARD; PRT; 354 AA.

AC Q9KPG7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNAc
DE transferase).
GN MURG OR VC2401.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OC NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
CC STRAIN=El Tor N16961 / Serotype O1;
CC MEDLINE=20406833; PubMed=10952301;
CC Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
CC Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
CC Gill S.R., Nelson K.B., Read T.D., Tettelin H., Richardson D.,
CC Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
CC McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
CC Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
CC Fraser C.M.;
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)/GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

DR TIGRFAMS; TIGR01133; murG; 1.
KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 355 AA; 38478 MW; 5B74CBE2BFB6574 CRC64;

Query Match 54.3%; Score 1021; DB 1; Length 355;
Best Local Similarity 57.5%; Pred. No. 1.3e-67;
Matches 204; Conservative 51; Mismatches 92; Indels 8; Gaps 3;

QY 7 KRLVMAGGTHGHHVPGGLAVAHLMAGQGVVRLGTADRMEDLVPHKGIEIDFIRISGL 66
DB 3 KKLILMAGGTHGHHVFPFAIAVAQELQKQWICWLTGKDRNEALVQYNIPIEFQISGL 62
QY 67 RKGKIKALIAAPLIRIFNAWQARAIMKAYKPDVVLGMGGVSGPGGLAWSIGIPVVLHE 126
DB 63 KKGVLALIKAPFTILKAVLQALNIIKKYRPDAVLGMGGVSGPGGIAARLNCVPLVHE 122
QY 127 QNGTAGLTKNLARIATKVMQAEFGAPNAEVGVNPTVTDVLALPQORLAGREG- 182
DB 123 QNAIAGLTVNLAKIAKRVLAQPTAFKAEVTGVPVKDLSLLOPAQRFKARATAPY 182
QY 183 PVRVLVGGSGGARILNQTTPQVAAKLGSVIIWHQSGKGSQSQSVQAYAEAGQPHKVT 242
DB 183 PLNLVGGSGGARILNQTTPQVAAKLGSVIIWHQSGKGNLRTISDVYKQADNV--SVT 240
QY 243 EFIDMAAAYAMADVVCRSGLTVSIIAAGLPALFVFPQHKDROQYVNLPLEKAGAA 302
DB 241 EFIDMAEAYNADLVICRSGLTVSIIAAGLPALFVFPYQHKDROQYVNLPLEKAGAA 300
QY 303 KIIQPOLSDVAVANTLAGW--SRETLTWAERARAASIDATERVANEVSRVAR 355
DB 301 IIVEQPOFTAEINLNILOPKROKLTETWAKATKATKAAQVAVETIIEYSK 355

RESULT 16
MURG_BUCAP STANDARD; PRT; 354 AA.
AC Q8X9T4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNac
DE transferase)
GN MURG OR BUS210.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tanas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria";
RL Science 296:2376-2379 (2002).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNac
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)/GlcNac (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNac- (1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE014097; AAM67773.1; -.
DR HAVAP; MF_00033; -; 1.
DR InterPro; IPR007235; Glyco_tran_28.C.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR006009; MurG.
DR Pfam; PF04101; Glyco_tran_28.C; 1.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR TIGRFAMS; TIGR01133; murG; 1.
DR Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 354 AA; 39558 MW; 848B5F42605F2852 CRC64;

Query Match 50.2%; Score 945; DB 1; Length 354;
Best Local Similarity 48.1%; Pred. No. 4.7e-62;
Matches 169; Conservative 79; Mismatches 103; Indels 0; Gaps 0;

QY 7 KRLVMAGGTHGHHVPGGLAVAHLMAGQGVVRLGTADRMEDLVPHKGIEIDFIRISGL 66
DB 4 KRILLAGSGGHHVFPGLTIAKHLIKKGDINNTGKIESEIIPKCNIKIHIKIQGL 63
QY 67 RKGKIKALIAAPLIRIFNAWQARAIMKAYKPDVVLGMGGVSGPGGLAAWSIGIPVVLHE 126
DB 64 RNSLKNLIMTPINVLNSYLQVRKLIKWIPIIILGMGGVSGPGGLAANSCKIPFILHE 123
QY 127 QNGTAGLTKNLARIATKVMQAEFGAPNAEVGVNPTVTDVLALPQORLAGREGPVV 186
DB 124 QNKIAGITNKLKISTKNQAFSGTLLNAEIVGNPIKNIIDIPPIKRFKNRKGPLRI 183
QY 187 LVVGGSGARILNQTTPQVAAKLGSVIIWHQSGKGSQSQSVQAYAEAGQPHKVTBEID 246
DB 184 LIIGSGQASIFNKILPKIPFLQEKALIIHQSGNDLQKTRKYYKISTYKHIVSFIK 243
QY 247 DMAAAYAMADVVCRSGLTVSIIAAGLPALFVFPQHKDROQYVNLPLEKAGAAKIE 306
DB 244 NIAEAYEWADIIISRSGLTVSEITVGLGAIFFYPHKDQYVNLAEADLENNGAAKIE 303
QY 307 QPOLSDVAVANTLAGWSRETLTWAERARAASIDATERVANEVSRVARAL 357
DB 304 QSMFTAELIKILNSLNREKLFIMAKKAYSIGIRNSTSKISKIHDVSNKI 354

RESULT 17
MURG_BUCAI STANDARD; PRT; 354 AA.
AC P57311;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNac
DE transferase).
GN MURG OR BU216.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS";
RL Nature 407:81-86 (2000).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNac
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-

```
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac (oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac (oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP001118; BAB12932.1; -.
CC HSSP: P17443; 1F0K.
CC HAMAP: MF_00033; -.
CC InterPro: IPR007235; Glyco_tran_28_C.
CC InterPro: IPR004276; Glyco_tran_28.
CC InterPro: IPR006009; MurG.
CC Pfam: PF04101; Glyco_tran_28_C; 1.
CC TIGRFAMs: TIGR01133; murG; 1.
CC TrEMBL: F040101; Glyco_tran_28_C; 1.
CC Transferase: Glycosyltransferase; Cell division; Cell wall; Membrane;
CC Peptidoglycan synthesis; Complete proteome.
CC KW Peptidoglycan synthesis; Complete proteome.
CC SQ SEQUENCE 354 AA; 39499 MW; 70F23A1294E6B3A6 CRC64;

Query Match 49.9%; Score 939; DB 1; Length 354;
Best Local Similarity 48.4%; Pred. No. 1.3e-61;
Matches 169; Conservative 80; Mismatches 100; Indels 0; Gaps 0;

QY 7 KRLVMAGTGGHVPGLAVAHLMQAQGVRLGTADRMEDALVPKHGIEIDFIRISGL 66
DB 4 KXIIIMAGSGGHVFPFGITARIYETKGLVNWLTGNTKNSIESRIIPKYGIKHYSIKGL 63
QY 67 RGKGIKALIAAPLRIFENAWROARAIMKAYKPDVVLGMGGVYSGPGLAAMSGLGIPVYLHE 126
DB 64 RNTSLKNLIISPIYLTRAYAYVKIITKWSPDVILGMGGVYSGPGLAAMSGLGIPVYLHE 123
QY 127 QNGIAGLTNKLARIATKNQAEPGAFNAEVGNVPTDVLALPLQORLAGREGPVV 186
DB 124 QNKIAGITNKLRSITRNQASPGVLNNAEVGNVPTDVLALPLQORLAGREGPVV 183
QY 187 LVVGSQGARILNOTMPOVAALGDSVVIHQSGKGSQSQSQVEQAYABAGOPQHKVTEFID 246
DB 184 LVIGSQSSILNRLPEVSEFLKEKILFWHTQNGYLEKTKKYNKLRNLQNLTSIFIK 243
QY 247 DMAAYAWADVVCVSGALTYSETAAGLPALEFVFFQHKDQQQVWALPLEKAGAAKIE 306
DB 244 NIASAYEADLTCISGALTYSISIVGLGAIFIPYPHKQQRHNAEDLELGAAKIID 303
QY 307 QPQLSVDAVNTLACWSRETLITWAERARAASIPDATERVANEVSRAV 355
DB 304 QSNLTKLVNLINSLDKDLFIMAKKAHSLGVDRDAIFNIENFINKISK 352

RESULT 18
MURG_CANBF
ID MURG_CANBF STANDARD; PRT; 360 AA.
AC Q7U346;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-pe-MurNAc-pentapeptide-UDP-GlcNAc
DE transferase).
GN MURG OR BFU142.
OS Candidatus Blochmannia floridanus.
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; ant endosymbionts; Candidatus Blochmannia.
OX NCBI_TaxID=203907;
RN SEQUENCE FROM N.A.
RX MEDLINE=22784745; PubMed=12886019;
RA Gil R., Silva F.J., Zientz E., Delmotte F., Gonzalez-Candelas F.,
RA Latorre A., Rausell C., Kamerbeek J., Gadau J., Hoelldobler B.,
RA van Ham R.C.H.J., Gross R., Moya A.;
RT "The genome sequence of Blochmannia floridanus: comparative analysis
RT of reduced genomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:9388-9393(2003).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAc-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac (oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac (oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: BX248584; CAD83663.1; -.
CC HAMAP: MF_00033; -.
CC InterPro: IPR004276; Glyco_tran_28.
CC InterPro: IPR007235; Glyco_tran_28_C.
CC InterPro: IPR006009; MurG.
CC Pfam: PF03033; Glyco_tran_28; 1.
CC Pfam: PF04101; Glyco_tran_28_C; 1.
CC TIGRFAMs: TIGR01133; murG; 1.
CC TrEMBL: F040101; Glyco_tran_28_C; 1.
CC Transferase: Glycosyltransferase; Cell division; Cell wall; Membrane;
CC Peptidoglycan synthesis; Complete proteome.
CC KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
CC SQ SEQUENCE 360 AA; 40579 MW; DB0BE4539AD04F46 CRC64;

Query Match 47.8%; Score 900; DB 1; Length 360;
Best Local Similarity 50.7%; Pred. No. 9.4e-59;
Matches 180; Conservative 64; Mismatches 107; Indels 4; Gaps 3;

QY 2 MSGQGKRLVMAGTGGHVPGLAVAHLMQAQGVRLGTADRMEDALVPKHGIEIDFI 61
DB 1 MNGSVQTIIMIIAGTGGHVPGLSVARYLMNHGYKVVMWIGSKDRIESELVPPYNDIKYI 60
QY 62 RISGLEKGI-KALIAAPLRIFENAWROARAIMKAYKPDVVLGMGGVYSGPGLAAMSGLGI 120
DB 61 CIOGLRGKXIYQKLIITLLFLIFAMYSQSPKIRCKWPKDIVLNGVYSGPSSLVAVLYGI 120
QY 121 PVLVHEQNGIAGLTNKLARIATKNQAEPGAFNAEVGNVPTDVLALPLQORLAGR 180
DB 121 PVIHQENRILNGLTNRVYVSFAKKIILQGFNTVNGAITSNPLRVYELISIPDPVHLEGR 180
QY 181 EGPVRVLVVGSGGARILNOTMPOVAALGDSVVIHQSGKGSQSQSQVEQAYAE--AGQPQ 238
DB 181 TGPIRVLVVGSGTGSFIFNKVPEVFGKLFGLKLIHQSGKKGKGFNDTIQAYKKLHNSN 240
QY 239 HKYTEFTIDMAAYAWADVVCVSGALTYSETAAGLPALEFVFFQ-HKDRQQVWALPLE 297
DB 241 YKVVPIDNWAHAYSADVVIISGALMYSISVGLPAIFVFPFNKDYQQVWNAFQLV 300
QY 298 KAGAAKIIIEQPQLSVDAVNTLACWSRETLITWAERARAASIPDATERVANEVSR 352
DB 301 KSGSALIIIEQERTSDYVSIIILGNMNRKVLNNAIILSKSLEMSNATQLVAQTVMR 355
```

```
RESULT 19
MURG_SHEON
ID MURG_SHEON STANDARD; PRT; 362 AA.
AC Q8CX35;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNAC-pentapeptide-UDP-GlcNAc GlcNAc
DE transferase).
DE MURG OR SO4219.
GN Shewanella oneidensis.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadales; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RC MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealeson K.H., Fraser C.M.;
RT Genomic sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RL Nat. Biotechnol. 20:1118-1123 (2002).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(Oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(Oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE015855; AAN57191.1; -
CC DR TIGR; SO4219; -
CC DR HAMAP; MF_00033; -; 1.
CC DR InterPro; IPR004276; Glyco_trans_28.
CC DR InterPro; IPR007235; Glyco_tran_28_C.
CC DR InterPro; IPR006009; MurG.
CC DR Pfam; PF03033; Glyco_transf_28; 1.
CC DR Pfam; PF04101; Glyco_tran_28_C; 1.
CC DR TIGRFAMs; TIGR01133; murG; 1.
CC DR Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
CC inner membrane; Peptidoglycan synthesis; Complete proteome.
CC SEQUENCE 362 AA; 38380 MW; 061728AE95578F8FF CRC64;
Query Match 47.2%; Score 887; DB 1; Length 362;
Best Local Similarity 52.5%; Pred. No. 8.4e-58;
Matches 186; Conservative 60; Mismatches 98; Indels 10; Gaps 6;
QY 2 MSGGKRLMNVAGTGGHVFPLAVAHLMQAQGVWGLGTADRMEDADLPKKGIEIDFI 61
DB 1 MTDAGKRLNVNAGTGGHVFPLAVAKYLAQQGVWGLGTADRMEDADLPKGVGPDIDFI 60
```

```
QY 62 RISLGRKGKIKALIAAPLRIPNARQARAIMKAYKPDVILMGVGVSPGGLAAWSLGP 121
DB 61 DIKVGVRGLVRLKLAAPFKVRSILQAKAVIAEFKPDVILMGVGFASPGGVAALAGVP 120
QY 122 VYLHEQNGIAGLTNKLARIATKYNOABPGAPP--NAEVGNVPVTRDVLPLPQORLAG 179
DB 121 LVLHEQNAIPGWTNKLRSIASQVLCAPKNTFTQVKAKVGVGNPIRRELIALLGGPFKQTA- 179
QY 180 REGPVRLVVGSGOGARILNQTMTPOVAAKLG--DSVLIWHQSGKSGSQSVQEAQVAGQP 237
DB 180 -DEARLUVLVVGSGLGAKVFNDLMEFVVAALSKQSIIVHHQVGNKLAGVKSAYCQQGQD 238
QY 238 QH-KVTEFIDDMAAYAWADVVCRSGALTYSEIAAAGLPALFVPFQHK-DRQYVWNAALP 295
DB 239 GGVVAEFDIDMEAYRWADVVLCRAGALTVELAAVGLPSILVYPYHVAVDHQTNRNAVQ 298
QY 296 LEKAGAAKIIIEQPOLSVDAVAN--TLAGSRETLLTVAERARAASIPDATEVA 347
DB 299 LVEAGAAFLPQAILLDVNVKLVSKQLLNDRAELRMGQRARDVAVLDATEQVA 352
RESULT 20
MURG_RALSO
ID MURG_RALSO STANDARD; PRT; 365 AA.
AC Q8XVI7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNAC-pentapeptide-UDP-GlcNAc GlcNAc
DE transferase).
DE MURG OR RSC2844 OR RS00261.
GN Ralstonia solanacearum (pseudomonas solanacearum).
OS Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RC MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Inebaut P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RL "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502 (2002).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(Oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(Oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL646072; CAD16551.1; -
CC HAMAP; MF_00033; -; 1.
```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 07:02:25 ; Search time 45 Seconds

(without alignments)

2552.192 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 MMSQGKRLVMAGGTGGHV.....RVANEVSRAVALEHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_todent:*

12: sp_virus:*

13: sp_vertibrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1793	95.3	355	16	Q83MM4 shigella fl
2	1021	54.3	355	16	Q7U336 haemophilus
3	900	47.8	360	16	Q7U346 candidatus
4	887	47.2	362	16	Q8CX35 shewanella
5	794	42.2	357	16	Q7W4B4 bordetella
6	793	42.2	357	16	Q7WFS2 bordetella
7	791	42.1	357	16	Q7VUQ3 bordetella
8	787	41.8	358	16	Q820X3 coxiella bu
9	764	40.6	357	16	Q82VS3 nitrosomona
10	717	38.1	356	16	Q87WV5 pseudomonas
11	683	36.3	359	16	Q88N76 pseudomonas
12	506	26.9	410	16	Q7VEP8 mycobacteri
13	496.5	26.4	379	16	Q8CY39 bruceella su
14	480.5	25.5	360	16	Q8FNU0 corynebacte
15	476	25.3	372	16	Q8A2S8 bacteroides
16	461.5	24.5	363	16	Q820F6 streptomyc

379	2	Q8GBA0	Q8GBA0 heliobacill
393	16	Q8CY50	Q8CY50 bifidobacte
356	16	Q83HK1	Q83HK1 tropheryma
356	16	Q820Y4	Q820Y4 tropheryma
363	16	Q88V81	Q88V81 lactobacill
311	2	Q9RQK0	Q9RQK0 caulobacter
361	16	Q7V466	Q7V466 prochloroco
418	16	Q7W3U6	Q7W3U6 deinoxococ
358	16	Q7U3U6	Q7U3U6 deinoxococ
366	16	Q89FU7	Q89FU7 bradyrhizob
357	16	Q85N5	Q85N5 fusobacteri
357	16	Q85N5	Q85N5 fusobacteri
357	16	Q7VDZ2	Q7VDZ2 prochloroco
357	16	Q8CUL4	Q8CUL4 oceanobacil
369	16	Q893R7	Q893R7 clostridium
431	16	Q9C9T8	Q9C9T8 arabidopsi
364	16	Q7V388	Q7V388 prochloroco
352	16	Q812T8	Q812T8 bacillus ce
352	16	Q81J6	Q81J6 bacillus an
358	16	Q820E0	Q820E0 chlamydophi
350	16	Q7U322	Q7U322 halicobacte
354	16	Q812Y1	Q812Y1 bacillus ce
358	16	Q8F4J1	Q8F4J1 leptospira
352	16	Q8DQ1	Q8DQ1 streptococc
653	10	Q8LQD0	Q8LQD0 oryza sativ
357	16	Q8CMW3	Q8CMW3 staphylococ
360	2	Q9FB02	Q9FB02 streptococc
358	16	Q8E6P0	Q8E6P0 streptococc
358	16	Q8CX15	Q8CX15 streptococc
361	16	Q8DVE2	Q8DVE2 streptococc
356	17	Q8YD0	Q8YD0 methanopyru
388	16	Q7TY01	Q7TY01 mycobacteri
420	16	Q33282	Q33282 mycobacteri
397	16	Q825U1	Q825U1 streptomyc
425	2	Q83W14	Q83W14 streptomyc
396	2	Q9RQ5	Q9RQ5 streptomyc
402	16	Q88GN0	Q88GN0 rhizobium l
392	2	Q8KNF2	Q8KNF2 micromonosp
392	16	Q49841	Q49841 mycobacteri
402	2	Q9F8U7	Q9F8U7 streptomyc
427	16	Q82VF2	Q82VF2 rhizobium m
388	2	Q9RPA1	Q9RPA1 streptomyc
382	2	Q9F2F9	Q9F2F9 streptomyc
346	16	Q92Y00	Q92Y00 rhizobium m
525	10	P93115	P93115 cucumis sat
112	2	O54027	O54027 porphyromon
433	16	P73403	P73403 synecocyst
522	10	Q9SM44	Q9SM44 spinacia ol
379	2	Q93H13	Q93H13 streptomyc
378	16	Q7WP06	Q7WP06 bordetella
378	16	Q7W1A2	Q7W1A2 bordetella
418	16	Q8PJG6	Q8PJG6 xanthomonas
378	16	Q7VUE9	Q7VUE9 bordetella
519	17	O27324	O27324 methanobact
376	2	Q8KND7	Q8KND7 micromonosp
535	10	Q9FZL3	Q9FZL3 nicotiana t
237	17	O27683	O27683 methanobact
390	2	Q9ZGB8	Q9ZGB8 streptomyc
530	10	Q9FZL4	Q9FZL4 glycine max
559	16	Q8H2U4	Q8H2U4 pseudomonas
388	16	Q8UCA6	Q8UCA6 agrobacteri
390	2	Q8GHC2	Q8GHC2 streptomyc
411	16	Q9RVF3	Q9RVF3 deinoxococ
444	16	Q8P450	Q8P450 xanthomonas
370	16	Q9PEZ9	Q9PEZ9 xylella fas
391	2	Q9RP99	Q9RP99 streptomyc
357	16	Q87PP1	Q87PP1 vibrio para
352	2	P74819	P74819 shingomona
428	2	O88003	O88003 bordetella
430	16	Q7WR12	Q7WR12 bordetella
430	16	Q7W234	Q7W234 bordetella
463	2	Q50458	Q50458 mycobacteri
360	16	Q92VH0	Q92VH0 rhizobium m
435	16	Q9CD88	Q9CD88 mycobacteri

90 116.5 6.2 438 16 Q9CD91 mycobacteri
91 116 6.2 465 10 Q9S193 arabidopsis
92 115.5 6.1 91 2 Q93SU1 chlorobium
93 115.5 6.1 390 2 Q9ALM8 saccharopol
94 115.5 6.1 428 16 P95134 mycobacteri
95 115.5 6.1 1049 16 Q89LT6 bradyrhizob
96 115 6.1 388 16 Q9WZ90 thermotoga
97 115 6.1 428 16 Q45374 bordetella
98 115 6.1 449 16 P95130 mycobacteri
99 115 6.1 449 16 Q7TXJ4 mycobacteri
100 115 6.1 1273 17 Q9YCA1 aeropyrum p

ALIGNMENTS

RESULT 1
Q83MN4 PRELIMINARY; PRT; 355 AA.
ID Q83MN4
AC Q83MN4;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR SF0087 OR S0089.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
Yu J.;
RA "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
Schwartz D.C., Blattner F.R.;
RA "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE015046; AAN41752.1; -
DR EMBL; AE016978; AAP15633.1; -
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR006009; Murg.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRFAMs; TIGR01133; murg; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 355 AA; 37812 MW; 73407776C2B1504C CRC64;

Query Match 95.3%; Score 1793; DB 16; Length 355;
Best Local Similarity 98.6%; Pred. No. 1.5e-119;
Matches 350; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 MSGCKRLVMAGGTGGHVPPLGLAVAHILMAQGWVWMLGTADREADLVPKHGIEIDFI 61
DB 1 MSGCKRLVMAGGTGGHVPPLGLAVAHILMAQGWVWMLGTADREADLVPKHGIEIDFI 60
QY 62 RISGLRGKIGIKALIAAPLIRFNARQARAIMKAYKDPDVLGMGGYVSGPGGLAAWSLIGIP 121
DB 61 RISGLRGKIGIKALIAAPLIRFNARQARAIMKAYKDPDVLGMGGYVSGPGGLAAWSLIGIP 120
QY 122 VVLEHONGIAGLTNKLARITATKMQAEPGAFNPAEVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLEHONGIAGLTNKLAKIATKMQAEPGAFNPAEVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPRVVLVVGSGQARILNQTMPQVAAKLGDSDVIWHQSGKGSQSQVEQAYAEAGQPQHKV 241
DB 181 GPRVVLVVGSGQARILNQTMPQVAAKLGDSDVIWHQSGKGSQSQVEQAYAEAGQPQHKV 240
QY 242 TEFIDDMAAAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDROQYNNALPLEKAGA 301
DB 241 TEFIDDMAAAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDROQYNNALPLEKAGA 300
QY 302 AKIIEOPOLSDVAVANTLAGHSRETLTMAERARAASIPDATERVANEVSRAA 356
DB 301 AKIIEOPOLSDVAVANTLAGHSRETLTMAERARAASIPDATERVANEVSRAA 355
RESULT 2
Q7U336 PRELIMINARY; PRT; 355 AA.
ID Q7U336
AC Q7U336;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR H00824.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=35000HP / ATCC 700724;
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT "The complete genome sequence of Haemophilus ducreyi";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017153; AAP95721.1; -
KW Transferase; Complete proteome.
SQ SEQUENCE 355 AA; 38478 MW; EB74CBE2BFBB6574 CRC64;

Query Match 54.3%; Score 1021; DB 16; Length 355;
Best Local Similarity 57.5%; Pred. No. 1.3e-64;
Matches 204; Conservative 51; Mismatches 92; Indels 8; Gaps 3;
QY 7 KRLVMAGGTGGHVPPLGLAVAHILMAQGWVWMLGTADREADLVPKHGIEIDFI 66
DB 3 KLLIMAGGTGGHVPPLGLAVAHILMAQGWVWMLGTADREADLVPKHGIEIDFI 62
QY 67 RGKGIKALIAAPLIRFNARQARAIMKAYKDPDVLGMGGYVSGPGGLAAWSLIGIP 126
DB 63 KKGKGLALIKAPFTTLKAVLQALNIIKKYRPDVLGMGGYVSGPGGLAAWSLIGIP 122
QY 127 QNGIAGLTNKLARITATKMQAEPGAFNPAEVGNPVRTDVLALPLPQORLAGRE 182
DB 123 QNAIAGLTNKLAKIATKMQAEPGAFNPAEVGNPVRTDVLALPLPQORLAGRE 182
QY 183 PVRVLVVGSGQARILNQTMPQVAAKLGDSDVIWHQSGKGSQSQVEQAYAEAGQPQHKV 242
DB 183 PLNVLVVGSGQARILNQTMPQVAAKLGDSDVIWHQSGKGSQSQVEQAYAEAGQPQHKV 240
QY 243 EFIDDMAAAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDROQYNNALPLEKAGA 302
DB 241 EFIDDMAAAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDROQYNNALPLEKAGA 300

```
QY 303 KIIIEPOLSVDAVANTAGW--SRETLTWAERARRASIPDATERVANEYSRVAR 355
DB 301 IIVEQPDFAENLNILOPLIKRQKTEWAIKATKATPAARVAEVIIVSK 355

RESULT 3
QY346 PRELIMINARY; PRT; 360 AA.
AC Q7U346;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.-)
GN MURG OR BFL142.

OS Candidatus Blochmannia floridanus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; ant endosymbionts; Candidatus Blochmannia.
OX NCBI_TaxID=203907;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22784745; PubMed=12886019;
RA Gil R., Silva F.J., Zientz E., Delmotte F., Gonzalez-Candela F.,
RA Latorre A., Rausell C., Kamerbeek J., Gadau J., Hoelldobler B.,
RA van Ham R.C.H.J., Gross R., Moya A.;
RA "The genome sequence of Blochmannia floridanus: comparative analysis
RT of reduced genomes";
RT Proc. Natl. Acad. Sci. U.S.A. 100:9388-9393(2003).
RL EMBL; BX248584; CAD8363.3; -.
RW Glycosyltransferase; Transferase; Complete proteome.
KW SEQUENCE 360 AA; 40579 MW; DB08B4539AD04F46 CRC64;

Query Match 47.8%; Score 900; DB 16; Length 360;
Best Local Similarity 50.7%; Pred. No. 5.4e-56;
Matches 180; Conservative 64; Mismatches 107; Indels 4; Gaps 3;

QY 2 MSGGQKLMWAGTGGHVPGLAVAHHLMAQGVQVWLGTADRMEDLVPKHGIEDFI 61
DB 1 MNQSVQTIMIAGTGHHIPGLSVARYLHMGYKVVYWGSKDRIESELVPEVNIIDKIY 60

QY 62 RISGLRGKGI-KALIAAPLRFNARQARIMKAYKPDVVLGMGVGSGGGLAAVSLGI 120
DB 61 CIOGLRGKVIKQKILTLFLFFMYQSFKILRCWKEDVILSMGVGSGPSSIVAMLYGI 120

QY 121 PVVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAEVVGNVPTDVLALPLPQORLAGR 180
DB 121 PVLIHEQNRINGLTNRYVSRFAKKILQGFNTVNGAITSNPLRYEILSIDPDVHRLEGR 180

QY 181 EGPVRLVVGSGQARILNQTMPQVAAKLGDSVLIWHQSGKSGSQSVQVQAYAE--AGQPQ 238
DB 181 TGPVRLVVGSGTSGFIFNKVPEVFGKLFKLLIWHQSGKSGKGFNDTIQAYKKLHCNSSN 240

QY 239 HKVTEFIDMAAAYAWADVVCVRSAGLTVEIAAAGLPALFVFPQ-HKDRQYQYVNALPFE 297
DB 241 YKVVPFDINNAHAYSWADVIISRSGLAMVSEISVGLPAIFVFNHYKQVQYVNAFQLV 300

QY 298 KAGAAKIIEQPOLSVDAVANTAGWSRETLTWAERARRASIPDATERVANEYSR 352
DB 301 KSGSALIIIEQRF*SDYVSIILGNWRKVLINMAILSKLSMSNATQLVAQTVMR 355

RESULT 4
QY346 PRELIMINARY; PRT; 362 AA.
AC Q8CX35;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR S04219.
```

```
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MR-1;
RC MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Mathe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imbraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feildblum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RT Nat. Biotechnol. 20:1118-1123(2002).
RL EMBL; AF015855; AAN57191.1; -.
DR TIGR; S04219; -.
DR GO; GO:0013866; C:inner membrane; IEA.
DR GO; GO:0016758; P:transferase activity, transferring hexosyl .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030253; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylgalactosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_trans_28.
DR InterPro; IPR006009; Murg.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_trans_28_C; 1.
DR TIGRFAMs; TIGR01133; murG; 1.
KW Complete proteome.
RW SEQUENCE 362 AA; 38380 MW; 061728AE95578FEF CRC64;

Query Match 47.2%; Score 887; DB 16; Length 362;
Best Local Similarity 52.5%; Pred. No. 4.6e-55;
Matches 186; Conservative 60; Mismatches 98; Indels 10; Gaps 6;

QY 2 MSGGQKLMWAGTGGHVPGLAVAHHLMAQGVQVWLGTADRMEDLVPKHGIEDFI 61
DB 1 MTDAGKRLVWAGTGGHVPGLAVAHHLMAQGVQVWLGTADRMEDLVPKHGIEDFI 60

QY 62 RISGLRGKGI-KALIAAPLRFNARQARIMKAYKPDVVLGMGVGSGGGLAAVSLGI 121
DB 61 DIKGVRLVVGSGQARILNQTMPQVAAKLG--DSVLIWHQSGKSGSQSVQVQAYAEAGQP 120

QY 122 VYLHEQNGIAGLTNKLARIATKVMQAEFGAP--NAEVVGNVPTDVLALPLPQORLAG 179
DB 121 LVLHEQNAIPGMNKLKSLASQVLCAPKNTFTQVAKVGVNPIRRELIALGGEPKQTA- 179

QY 180 RSGPVRVLVVGSGQARILNQTMPQVAAKLG--DSVLIWHQSGKSGSQSVQVQAYAEAGQP 237
DB 180 -DEALKVLVVGSGLGAKFVNDLMEPVVAALSKQSQITVWHQVGNKDLACVSAVQQQQQD 238

QY 238 QR-KYTEFIDMAAAYAWADVVCVRSAGLTVEIAAAGLPALFVFPQHK-DROQYVNALP 295
DB 239 GGVNVAEPIDDMEAAAYRWADVVLCEAGALTVELAAVGLPSILVFPYHVAVDHDTNRNAV 298

QY 296 LKAGAAKIIEQPOLSVDAVAN--TLAGWSRETLTWAERARRASIPDATERVA 347
DB 299 LVEAGAAFLPQAILDVNKLVLQQLANDRAELARMGQRRADVAVLDAATSOVA 352

RESULT 5
QY346 PRELIMINARY; PRT; 357 AA.
AC Q7W4B4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (BC 2.4.1.-).
GN MURG OR BP3752.
```


OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
CX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baskin M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Chillingworth T., Collins M., Cronin A., Hauser H., Holroyd S., Jagels K.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640434; CAE39035.1; --
KW Glycosyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 357 AA; 37893 MW; 48E5E646F5993C2 CRC64;

Query Match 42.2%; Score 794; DB 16; Length 357;
Best Local Similarity 48.3%; Pred. No. 1.9e-48;
Matches 169; Conservative 56; Mismatches 121; Indels 4; Gaps 3;

QY 9 LMWAGTGGHVFGLAVAHILMAQGVQVWLTGTRMEADLVPKHGIEDFIRISLGRG 68
DB 6 ILIMAGTGGHIMFGLAVAEVLRERGRVRLWLNPNKMGRLVPPRGIELVPLRFQVGRG 65
QY 69 KGKALIAAPLRIFNAWRQARIMKAYKPDVLMGCVGSGPGLAASLIGIPVVLHEQN 128
DB 66 RGAALALKLPFLARACACQAWRLADIRPDVLMGCVGVAFFPGVMAALRETPLVHEQN 125
QY 129 GIAGLTNKLARIATKYNQAEPPGAFNAEVVGNPVRTDVLALPLPQORLAGRGVRLV 188
DB 126 AVAGTANRLARLARRVLSGPPGLPRGELGNPVRADLCALPEPAERYAGRSALRLV 185
QY 189 VGGSGGARIINQTMPOVAALGLGDSV--IIWHOSGKGSQQSVEQAYABAGQPHKVTFTFD 246
DB 186 VGGSLGAHALNTTVFQALALLPEQARQVQVHAGEQHLPALQOAYAQAG-VQADCFRFD 244
QY 247 DMAAYAWADVVCRCGALTVEITAAAGLPALFPFQHK-DRQYWNALPLEKAGAKII 305
DB 245 DMAGAMAQADLLICRAGAMTVSEVAAAGVAALFVFPFHAIIDHQTANARFLSDAQAAWLQ 304
QY 306 EQPOLSDVAVANTLAGSRETLTMAERARAASIPDATERVANEVSRRVAR 355
DB 305 PQALITPQWLAQWLQGTQRCLOQAVAGRTALPRAAAHIADVCEQAAR 354

RESULT 6
QYWF52 PRELIMINARY; PRT; 357 AA.
ID QYWF52
AC QYWF52
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (EC 2.4.1.-).
GN MURG OR B34198.
OS Bordetella bronchiseptica (Alcaligenes bronchiseptica).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,

RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baskin M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640434; CAE34562.1; --
KW Glycosyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 357 AA; 37849 MW; BC5D8687E30F48AF CRC64;

Query Match 42.2%; Score 793; DB 16; Length 357;
Best Local Similarity 48.3%; Pred. No. 2.2e-48;
Matches 169; Conservative 55; Mismatches 122; Indels 4; Gaps 3;

QY 9 LMWAGTGGHVFGLAVAHILMAQGVQVWLTGTRMEADLVPKHGIEDFIRISLGRG 68
DB 6 ILIMAGTGGHIMFGLAVAEVLRERGRVRLWLNPNKMGRLVPPRGIELVPLRFQVGRG 65
QY 69 KGKALIAAPLRIFNAWRQARIMKAYKPDVLMGCVGSGPGLAASLIGIPVVLHEQN 128
DB 66 RGAALALKLPFLARACACQAWRLADIRPDVLMGCVGVAFFPGVMAALRETPLVHEQN 125
QY 129 GIAGLTNKLARIATKYNQAEPPGAFNAEVVGNPVRTDVLALPLPQORLAGRGVRLV 188
DB 126 AVAGTANRLARLARRVLSGPPGLPRGELGNPVRADLCALPEPAERYAGRSALRLV 185
QY 189 VGGSGGARIINQTMPOVAALGLGDSV--IIWHOSGKGSQQSVEQAYABAGQPHKVTFTFD 246
DB 186 VGGSLGAHALNTTVFQALALLPEQARQVQVHAGEQHLPALQOAYAQAG-VQADCFRFD 244
QY 247 DMAAYAWADVVCRCGALTVEITAAAGLPALFPFQHK-DRQYWNALPLEKAGAKII 305
DB 245 DMAGAMAQADLLICRAGAMTVSEVAAAGVAALFVFPFHAIIDHQTANARFLSDAQAAWLQ 304
QY 306 EQPOLSDVAVANTLAGSRETLTMAERARAASIPDATERVANEVSRRVAR 355
DB 305 PQALITPQWLAQWLQGTQRCLOQAVAGRTALPRAAAHIADVCEQAAR 354

RESULT 7
QYVUQ3 PRELIMINARY; PRT; 357 AA.
ID QYVUQ3
AC QYVUQ3
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (EC 2.4.1.-).
GN MURG OR B31023.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Baskin M., Cronin A., Davis P., Doggett J.,
RA Chillingworth T., Collins M., Cronin A., Hauser H., Holroyd S., Jagels K.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";

RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640420; CAE43294.1; --
KW Glycosyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 357 AA; 37907 MW; 2730305E9A9CB893 CRC64;
Query Match 42.1%; Score 791; DB 16; Length 357;
Best Local Similarity 48.3%; Pred. No. 3e-48;
Matches 169; Conservative 55; Mismatches 122; Indels 4; Gaps 3;
QY 9 LNMVAGTGGHVFPGCLAVAHMLMAQGVQVRLGTADRMADLVPHKGIETDFIRISGLRG 68
DQ :::
Db 6 ILIVAGTGGHIMPLGLVAEVLREGRVNLGLNPDKNMEGLVPPRGIELVPLRFQVRG 65
QY 69 KGKALIAAPLRIENAFQARAIKAVKPDVVLGMGGVSPGGLAAWSLIGIPVVLHEQN 128
DQ :::
Db 66 RGAALLKLPPELLARCAQAVRLADIPDVVLGMGGVAFPGVMAALRTPLVVEQN 125
QY 129 GIAGLTNKLARIATKVMQAEFGAPFNAEVVGNVPTDVLALPLPOORLAGREGPVVLV 188
DQ :::
Db 126 AVAGTANRWLARLARRVLSGPGVLPRLGELGNFVRADLCALPEPAERYAGRSGLRV 185
QY 189 VGSOGARILNCTMPQVAAKLGDSV--IITHSGKSGSQSVQVAYAGAQPHKVTETID 246
DQ :::
Db 186 VGSIGALHNTTVPQALLALPEQARPQVHVQAGEHLPALQAYAQAG-VQACRAFD 244
QY 247 DMAAAYAWADVVCVRSGLTIVSEIAAAGLPALFVPFQHK--DROQVYNALPLEKAGAAKII 305
DQ :::
Db 245 DWADAVAQADLLICRAGAMTVSEVAAAGVAALFVPPFPHAIIDHQTANARFLSDQAALW 304
QY 306 EPQQLSVDAVANTLAGSRETLTVAERARAASIPDATERVANEVSRAV 355
DQ :::
Db 305 POASLTPOWLAQWLGORTQELQAVAGARTHALPRAAAHIAADVCEQAR 354

RESULT 8
Q820X3 PRELIMINARY; PRT; 358 AA.
ID Q820X3
AC Q820X3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE Pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR CB00135.
OS Coccidia
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coccidiaceae; Coccidia.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I / RSA 493;
RX MEDLINE=22608657; PubMed=12704232;
RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Nelson W.C., Ward N.L., Tettelin H., Daviden T.M., Beanan M.J.,
RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
RA Thompson H.A., Samuel J.B., Fraser C.M., Heidelberg J.F.;
RT "Complete genome sequence of the Q-fever pathogen, Coccidia
burnetii";
RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
RL EMBL; AB016960; AAO9699.1; --
DR TIGR; CB00135; --
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_trans_28_C.
DR InterPro; IPR006009; Murg.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_trans_28_C; 1.
DR TIGR; TIGR01133; murG; 1.
KW Glycosyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 357 AA; 38390 MW; F4E1828EF0ED80FB CRC64;

Query Match 40.6%; Score 764; DB 16; Length 357;
Best Local Similarity 46.0%; Pred. No. 2.5e-46;

DR TIGR; TIGR01133; murG; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 358 AA; 39973 MW; 5BF7E97212E90AB6 CRC64;
Query Match 41.8%; Score 787; DB 16; Length 358;
Best Local Similarity 47.4%; Pred. No. 5.9e-48;
Matches 167; Conservative 64; Mismatches 101; Indels 20; Gaps 10;
QY 8 LNMVAGTGGHVFPGCLAVAHMLMAQGVQVRLGTADRMADLVPHKGIETDFIRISGLR 67
DQ :::
Db 3 RILLIAGTGGHIFPALAVARELREQVDVQVLMGVKGLLEKLP--DSPPLHLIQKAPR 61
QY 68 GK-GIKALIAAPLRIENAFQARAIKAVKPDVVLGMGGVSPGGLAAWSLIGIPVVLHE 126
DQ :::
Db 62 GRGLQQLL-NELRLRAVFOAYRIIRPKPQVILGMGGVAGPGGLAAWITRTPLIHE 120
QY 127 QNGIAGLTNKLARIATKVMQAEFGAPF--NAEVV--GNPVRTDVLALPLPOORLAGREGP 183
DQ :::
Db 121 QNSIPGLTNRVLAKMAKFLQGPDTFFQNRKVIITGNFVTELVMPLPVRLAARRGP 180
QY 184 VRVLVGGSGGARILNCTMPQVAAKLI--GDSVIIHOSKSGSQSVQVAYAGAQPHKV 241
DQ :::
Db 181 LRLVLGSGSGARSINQKMLAALSSYPSEETIAVWHQTGQORDFEFTQKEY--EKIKI 239
QY 242 TETIDMAAAYAWADVVCVRSGLTIVSEIAAAGLPALFVPFQHK--DROQVYNALPLEKAG 300
DQ :::
Db 240 DNFISDMAGAYWADLVVCRAGALTIVCEIASVGVASIFIPYPAVDNHQFHNAFLEQAG 299
QY 301 AAKIIEBPQLSVDAVANTLAGW-----SRETLTVAERARAASIPDATERV 346
DQ :::
Db 300 AAIIEESILT---ETDLRWFEQAPQDRDLLTMAENARKLAKPEAVQVRV 347
RESULT 9
Q82VS3 PRELIMINARY; PRT; 357 AA.
ID Q82VS3
AC Q82VS3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyltransferase family 28 (SC 2.4.1.-).
GN MURG OR NE0991.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Arciero D.M., Hommes N.G., Whitaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
obligate chemolithoautotroph Nitrosomonas europaea";
RL J. Bacteriol. 185:2759-2773(2003).
RL EMBL; BX321859; CAD84902.1; --
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_trans_28_C.
DR InterPro; IPR006009; Murg.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_trans_28_C; 1.
DR TIGR; TIGR01133; murG; 1.
KW Glycosyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 357 AA; 38390 MW; F4E1828EF0ED80FB CRC64;

Matches 162; Conservative 68; Mismatches 114; Indels 8; Gaps 4;

QY 10 MYMAGTGGHVFPGGLAVAHLMAGQWVRWLGTADRMADLVPKHGIEIDFIRISGLRGK 69

DB 1 MIMAGTGGHVFPGGLAVARSWQANGWRIVWLGTENGWEALVPQHGSFIELINFSGLGK 60

QY 70 GIKALIAAPLRFINAWQARAIKAYKPDVVVLGMGGYVSPGGLAAMSIGIPVVLHONG 129

DB 61 KLSYVLLLPWLAQACWQSFILRRQOPQVVLGGYYPALPGGIVALLKGLKELLHEQNR 120

QY 130 TAGLTKNKLARIATKWAQABGAPPNAE---VVGPNVRTDLALPLPOORLAGREGPVR 185

DB 121 TAGLTKNLAKIADRIALAPALTSFENTRTVTPVTEIARLPSPPEARVAHRTGNLH 180

QY 186 VLVVGGSGARILNQTPQVAALGDS--VIIHQGKGSGQSVQOAYAEAGQPOHKYTE 243

DB 181 ILVVGSLGAQVNLTVLPQALSMIPEDQRPVVTHQSKAHLDAQAYADHGVTGNLVA- 239

QY 244 FIDDMAAYANADVVCBSGALTSETAARAGLPALFVPOHK--DROCYNNALPLEKAGAA 302

DB 240 FIENWAHYQDCDLVICRAGALTISELAAGVAGVILIPYVDDHQTANARFLSDYQAA 299

QY 303 KIIEPQSLSDAVANTLAGSRETLTMAERARAASIPDATERVANEVSERVA 354

DB 300 VLWQSELTAASLAQMLTCSRAQLQSMATHARALAMPEAAQTVAZACQOLS 351

RESULT 10

Q8/WY5 PRELIMINARY; PRT; 356 AA.

ID Q87WY5

AC Q87WY5;

DT 01-JUN-2003 (TREMBLrel. 24, Created)

DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide)

DE DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.

GN MURG OR PSPT04408.

OS Pseudomonas syringae (pv. tomato).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_taxid=323;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=DC3000;

RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,

RA Barry K., Utterback T., Van Aken S., Feldblyum T., Winn M.,

RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,

RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidson T.,

RA White O., Fraser C., Collier A.;

RT "Complete sequence of *Pseudomonas syringae*."

RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AEO16871; AAO57857.1; -

DR TIGR; PSPT04408; -

DR GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.

DR GO; GO:0005975; P:carboxylate metabolism; IEA.

DR GO; GO:0030259; P:lipid glycosylation; IEA.

DR InterPro; IPR004276; Glyco.trans.28.

DR InterPro; IPR007235; Glyco.trans.28.C.

DR Pfam; PF03033; Glyco.transf.28; 1.

DR Pfam; PF04101; Glyco.trans.28.C; 1.

DR Complete proteome.

QW SEQUENCE 356 AA; 3798 MW; 58956CE06DFC52C CRC64;

Query Match 38.1%; Score 717; DB 16; Length 356;

Best Local Similarity 44.1%; Pred. No. 5.6e-43;

Matches 159; Conservative 63; Mismatches 117; Indels 20; Gaps 9

QY 9 LMYMAGTGGHVFPGGLAVAHLMAGQWVRWLGTADRMADLVPKHGIEIDFIRISGLRG 68

DB 5 VLMAGTGGHVFPGGLAVAHLMAGQWVRWLGTADRMADLVPKHGIEIDFIRISGLRG 64

QY 69 GKIKALIAAPLRFINAWQARAIKAYKPDVVVLGMGGYVSPGGLAAMSIGIPVVLHON 128

65 KRRLSLKAPMLIKALMQARKVVRQVPCVVGFGVGYTGPGLAARLAGVPLIIHEQN 120

129 GIAGLTNKWLARIATKYVQMAEPGAF---PNAEVGVNPNRTDVLALPLPQORLAGREGPVR 185

125 AVAGTANESLASFAFRSCEAPPNTFAASAKRRITGTENFVRVE-LFLETPTQALAGRK--AR 181

186 VLAVGSGOGARILNQTPQVAAKLGDSV--LIHQSGKSGQSQVQAYABAGQPOHKVTE 243

182 LVLVGGSLGAPLNKLGLPDALAAQAPQDIQEVFHQSGKNHDATARYRNVG-VEAQVAP 240

244 FIDDMAAAYAWDVVVCRSGALTYSIEIAAAGLPALFVPFHQK-DRQYWNALPLEKAGAA 302

241 FIONMAQVSWADLWCFSAGALTISLAAAGLPALLPLPHADDDHQSRYADYLAEGAA 300

303 KIIQEPOLSVDAVNTLAGMSRETL-----TWAEARAASIPDATERVANEVSVA 354

301 FVM--PQATTGAA--EWAARLUKEVLMQPEQLNSWARTARSIAKPEDATNTVVVCVEVA 354

RESULT 11

Q88N76 PRELIMINARY; PRT; 359 AA.

AC Q88N76; 24, Created)

DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)

DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.

GN MORG OR PP1337.

OS Pseudomonas putida (strain KT2440).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxId=160488;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22423050; PubMed=12534463;

RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,

RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,

RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,

RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,

RA Chris Lee P., Rizzoapple E., Scanlan D., Tran K., Mozzez A.,

RA Unterbach T., Holczapfel K., Kosack D., Moestl D., Wedler H.,

RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,

RA Kiewitz C., Eisen J., Timmis K.N., Duescherhoef A., Tuemmler B.,

RA Fraser C.M.;

RT "Complete genome sequence and comparative analysis of the

RT metabolically versatile Pseudomonas putida KT2440.;"

EL Environ. Microbiol. 4:799-808(2002).

DR EMBL; AE016779; AAN66960.1; "

DR TIGR; PP1337; "

DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR GO; GO:0030259; P:lipid glycosylation; IEA.

DR InterPro; IPR004276; Glyco_tran_28.

DR InterPro; IPR007235; Glyco_tran_28 C.

DR Pfam; PF03033; Glyco_transf_28; 1.

DR Pfam; PF04101; Glyco_tran_28_C; 1.

KW Complete proteome.

SQ SEQUENCE 359 AA; 38127 MW; CF71810DFD8A4FC7 CRC64;

Query Match 36.3%; Score 683; DB 16; Length 359;

Best Local Similarity 43.2%; Pred. No. 1.5e-40;

Matches 156; Conservative 60; Mismatches 133; Indels 12; Gaps

QY 2 MSQSGKELVMAGGTCGHVFGPLGVAHLMQAQGWVRLGTADRMADLVKPHGIEIDFI 61

DB 1 MAAGEKNVLLIMAGTGCGHVFPPALACAREFQAGSYVHVLGTGPRGIENELVQAGLPHLI 60

QY 62 RISGLRGKGIKALIAPLRIFNWRQARIMAKYKPDVVLVGGYVSGPGGLAWSIGIP 121

DB 61 QVSGLRGKGLSLLKAPFTLVKAVLQARRIRIQLKPCVCLGFGGVTGPGVAAARLQGV 120

QY 122 VULHEQNGINGLTKWLARIATKYVQMAEPGAFNAA---VVGPNFRTDVLALPLPQORLA 176

```
Db 121 LVIHQNARACTANRLVPLSARVCEAFFNTFEASDKRRTTGNVPRPELF---MDAQRTP 177
Qy 179 GREGFVRVLVGGSCGAKILNQTWPQVAAKLGDSV--IIWHOSGKSQSQSVQAYAEAGQ 236
Db 178 LGERARLLVGGSLGAEPLNKLFPKALSEVPAALRPEVPHOAGKQAPITABRYHEAGV 237
Qy 237 POHKVTEFIDMAAYAWADVVCRSGALTIVSEIAAAGLPLALFPVPOHK-DRQOYNWALP 295
Db 238 AA-QVEPFIDMAQAYGWADLVVCRAGALTIVSELAAAGLPSMLVPLPHADDDHQTNAQY 296
Qy 296 LEKAGAAKIIQPOLSVDAVANTL--AGWSRETLTWAERARAASIPDATERVANEVSRV 353
Db 297 LAREGAFLPQATTGAQALAERLNEVLMOPEKLVNMGATARRLAKPAATSTVVDICLEV 356
Qy 354 A 354
Db 357 A 357

RESULT 12
QVPEP8 PRELIMINARY; PRT; 410 AA.
AC Q7VEP8;
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DE UPD-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol-N-acetylglucosamine transferase Murg (EC 2.4.1.-)
GN MURG OR M2177C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12789972;
RA Garnier T., Eglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondest S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248341; CAD97030.1; --
DR Transferase; Glycosyltransferase; Complete proteome.
SQ SEQUENCE 410 AA; 41858 MW; 347DB45D34A4890B CRC64;

Query Match 26.9%; Score 506; DB 16; Length 410;
Best Local Similarity 37.8%; Pred. No. 6.9e-28;
Matches 143; Conservative 55; Mismatches 148; Indels 32; Gaps 10;

Qy 3 SQGKRLVMAGGTGGHVPGLVAHLMAGQVQR--WLGTADMEADLVPKHGIEIDF 60
Db 32 SADSLSVLAGGTAGHVEPAMAVADALVALDPRVITAGLTGPRGLETRLPVQGYHLEL 91
Qy 61 IRLSGLRGKIGKIALIAPLRFINAWROARIMKAYKPDVLMGVTGSGPGLAWSL-- 118
Db 92 ITAVPMRPGGDLARLPSRVVRAVREARDVLDVDDVAVVGVFGYVLPALYLAARGLPL 151
Qy 119 -----GTPVLVHEQNGIAGLTNKLWARIATKVMQAEF-GAPPNAEVVGNFVRTDVLALP 171
Db 152 PPRRRRFPVVIHEANARAGLANRYGAHTADRVLSAVPDSGLRRAEVVGVVPRASIALD 211
Qy 172 LFOQLAGR-----EGPVRVLV--GSGOARILNQTWPQVAAKLGDS-VIIWHOSGKSQ 224
Db 212 RAVLRAEAFARHFPPDDARVLLVFGSGQASVLSNRAVSGAAADLAAAGCVLH--AHGPQ 269
Qy 225 QVGEQAYAEAGOPQKVTETEFIDMAAAVADVVVCRSGALTIVSEIAAAGLPLALFPVPOH 284
Db 270 NVLELRRRAQGPVAVVAYFYLDRMELAYAADLVICRAGMTVAEVSAGVLPALVPLPI 329
```

```
Qy 285 KDRQOYNWALPLEKAGAAKIIQPOLSVDAVANTLWAGWSRETLTWAERARAASIPDATE 344
Db 330 GNGEQLNALPVNAGGVMVADALTPELVARQVAG-----LLT--DPARLAANTAAAA 382
Qy 345 RV-----ANEVSRVARAL 357
Db 383 RVGHRDAAGQVAAALAV 400

RESULT 13
Q8CY39 PRELIMINARY; PRT; 379 AA.
AC Q8CY39;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE UDP-N-acetylglucosamine--N-acetylmuramyl-(Pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR BR1431.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=2247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tetelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR EMBL; AE014438; AAN30344.1; --
DR TIGR; BR1431; --
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco trans 28.
DR InterPro; IPR007235; Glyco tran_28_C.
DR InterPro; IPR006009; Murg_28_C.
DR Pfam; PF03033; Glyco.transf.28; 1.
DR Pfam; PF04101; Glyco.tran_28_C; 1.
DR TIGRFAMs; TIGR01133; murg; 1.
KW Complete proteome.
SQ SEQUENCE 379 AA; 40236 MW; 705EE98D96F30177 CRC64;

Query Match 26.4%; Score 496.5; DB 16; Length 379;
Best Local Similarity 37.0%; Pred. No. 3e-27;
Matches 134; Conservative 53; Mismatches 154; Indels 21; Gaps 10;

Qy 2 MSQGKRLVMAGGTGGHVPGLVAHLMAGQVQRWLGT---ADMEADLVPKHGIEI 58
Db 4 LANQGV-IVLAGGTGGHLPFAEALAEHLARGWDVH-LATDARAQRFVGAFAQDH--V 58
Qy 59 DFTIRISGLRGKIGKIALIAPLRFINAWROARIMKAYKPDVLMGVTGSGPGLAWSL 118
Db 59 HVIRSATIAGRNPVALLKTFWSLMOQNLDRLKFLRELKPLVVGFGYVTPLPYLAASN 118
Qy 119 GIPVVLVHEQNGIAGLTNKLWLA---RIATKVMQAEFGAPPNAEVV-GNPFVRTDVLALP 173
Db 119 GIPVVLVHEQNGIAGLTNKLWLA---RIATKVMQAEFGAPPNAEVV-GNPFVRTDVLALP 178
Qy 174 QOQLAGRGEPVRLVVGSGOARILNQTWPQVAAKLGDS-----SVIIWHOSGKSQSVQ 229
Db 179 PYTPAGKDRDRFLLVFGSGQAGFFSCAIPAAVALLPEHERARLLITQARKEDASAQ 238
Qy 230 AYAEAGOPQKVTETEFIDMAAAVADVVVCRSGALTIVSEIAAAGLPLALFPVPOHK-DRQ 288
```

```
Db 239 AYKLGVPA-DVAPFFNDMPARMADAHFVIRGASTVSEITVIGFPMVLVPPHDLHD 297
Qy 289 QYNALPLEKAGAAKITEQPOLSDAVANTL--AGWSRETLTMAERARAASIPDAT 346
Db 298 QANAAALAAAGAEVVRQADLPQRLAEMQLQSAEMPERLEQQAARAKSVGKPDARLL 357
Qy 347 AN 348
Db 358 AD 359

RESULT 14
Q8ENU0
ID Q8ENU0 PRELIMINARY; PRT; 360 AA.
AC Q8ENU0
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE Peptidoglycan biosynthesis protein Murg.
GN MURG OR C82053.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005221; BAC1863.1; -.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_trans_28.
DR InterPro; IPR001092; HLH_bas.
DR InterPro; IPR006009; Murg.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_transf_28_C; 1.
DR TIGRfams; TIGR01133; murg_1; 1.
DR PROSITE; PS00038; HLH_1; 1.
KW Complete proteome.
SQ SEQUENCE 360 AA; 36543 MW; 1B37ACB96048E0D CRC64;

Query Match 25.5%; Score 480.5; DB 16; Length 360;
Best Local Similarity 35.6%; Pred. No. 3.8e-26;
Matches 132; Conservative 55; Mismatches 137; Indels 47; Gaps 9;

Qy 9 LMVWAGTGGHVPFGLAVAHHL--AQGVQVRLGTADRMADLVPKHGIEIDFIRISGLR 67
Db 10 VVWAGGTAGHIEPALAVAEALRDYCATVTALGTARGLETSLVDPDGLRIEIPVVP 69
Qy 68 KGKIKALIAAPLRFNMAWQARAIMKAYKPDVVLGMGVGPGGLAANSLGIPVVLHEQ 127
Db 70 RTPNLDLVKLPFRVAKSLRQARVLTGAGHVVGGYVSAPAYLAARSLGIPFFVHEA 129
Qy 128 NGIAGLNKWLARATKVMQAEPAFFNAEVNVPVTRTDVLPQRLAGREGPV---184
Db 130 NARAGMANKLGLVRLGGVGLNATENSGMPGVVGIPIR-----RELAGEDATAAE 179
Qy 185 -----RVLVVGGSGARILNQTMPQVAKLGD-----SVIIHQSGKGSQQSV 227
Db 180 RGRQTWGLEADRPVTLVTTGSGQSVNSA---VAGALDDLGLAGIQVLHAVGK--RNSL 234
Qy 228 EQAYEAGOPQHKYTEITDDMAAAYANADVVCVRSGLTVSEIAAGLPAFPVPOHKOR 287
Db 235 PTA-----RPGYVVPFIEDMQAAYAVADLVICVSGANTVAETASGIPAIYVPLPHNG 289
```

RESULT 15

```
Q8A258
ID Q8A258 PRELIMINARY; PRT; 372 AA.
AC Q8A258
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN BT3448.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL Science 299:2074-2076(2003).
DR EMBL; AB016940; AA078554.1; -.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_trans_28.
DR InterPro; IPR006009; Murg.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_transf_28_C; 1.
DR TIGRfams; TIGR01133; murg_1; 1.
DR Complete proteome.
SQ SEQUENCE 372 AA; 40242 MW; 2229D02DFF85921A CRC64;

Query Match 25.3%; Score 476; DB 16; Length 372;
Best Local Similarity 33.6%; Pred. No. 8.3e-26;
Matches 124; Conservative 78; Mismatches 143; Indels 24; Gaps 11;

Qy 8 LMVWAGTGGHVPFGLAVAHHL--MAQGVQVRLGTADRMADLVPKHGIEIDFIRISG 65
Db 6 RIISGGTGGHIFPAVSIANAIIELRPDAKILFVGAERMEMQVRPDAGYKILGLPIAG 65
Qy 66 LRKGKIKALIAAPLRFNMAWQARAIMKAYKPDVVLGMGVGPGGLAANSLGIPVVLH 125
Db 66 PDRKHLKMNVSILKLAESQWKAESILKNRFPQVAVGVGGVAGTTLKTAGMVGPTLIQ 125
Qy 126 FONGIAGTNTKWLARIATKVQAEPGA---PPNAEVV--GNPVTDLVALPLPQRLAGR 180
Db 126 EONSAGYVTKLLAQKAKAICVAYDGMKEKFPADKIINTGNPVRQN-LTKDMPKGAALR 184
Qy 181 EGPVR-----VLVVGSGQARILNQTMPQVAKLGD-----GDSVIIHQSGKGSQQSV 232
Db 185 SFNLPQDKTILVGGSLGARTINNTLTAALATIKENNDIQFIW-QTGKYVYPQVTEAVR 243
Qy 233 EAGQ-PQHKVTEITDDMAAAYANADVVCVRSGLTVSEIAAGLPAFPVPOHK-KDRQY 290
Db 244 AAGELPNLYTDFDKDMAAAYASDLVISRAGSISEFCILLHKPVLVLPSPNVAEDHQT 303
Qy 291 WNALPLEKAGAAKITEQPOLSDAVANTL--AGWSRETLTMAERARAASIPDAT 347
```

Db 304 KNALALVDKQAIYVYKDEAEAKLMDVALNTVA--DORKLXELSENTAKLALPDSABIIA 361
Qy 348 NEVSRVAPA 356
Db 362 QEVKLAEA 370

RESULT 16
Q820F6 PRELIMINARY; PRT; 363 AA.
AC Q820F6;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)pyrophosphoryl-
DE undecaprenol N-acetylglucosamine transferase.
GN MURG OR SAV6122.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMS 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Oonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMS 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AF005045; BAC73833.1; -.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030255; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_tran_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR006009; MurG.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRfam; TIGR01133; murG; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 363 AA; 38524 MW; B085F2493277597C CRC64;

Query Match 24.5%; Score 461.5; DB 16; Length 363;
Best Local Similarity 34.9%; Pred. No. 8.6e-25;
Matches 130; Conservative 62; Mismatches 149; Indels 31; Gaps 10;

Qy 9 LMVAGGTGGHVPGLVAHHLMAQGV--RWLGTADRMADLVPHKGIIDFIRISGL 66
Db 3 VVLAGGTGAGHIEPALADALRRDDPTVGITGTERGLTRLPVPERGYDLALIPAVPL 62
Qy 67 RKGKIKALIAAFLRTFNARQARATMKAYKDPVLMGVGVYSGPGLAAWSLGI;PVVLHE 126
Db 63 PKPTPELITVPCRIGRTIKAEQILERTKADAVVFGYVALGYLAANKELGVFIHIE 122
Qy 127 QNGIAGLTNKLARIATKVMQAEPGA--FNFVWGNVPRVTDVLAIP-----LPQORLAGRE 181
Db 123 ANARPGLANKIGSRVAAQAVVSTPDSKLRGARYIGIPLRRSIATIDRAAVRFEARAAFL 182

Qy 182 GP--VRLVTVGSGQARILNQTMPOVAKLGDSVI-IWHQSGKSGSQSVQVQYAEAGQ 238
Db 183 DNLPLTLVSGSGQARRLNEVVOQVAPYLOAGIQLH--AVGPKNEPQVHMPGMP 240
Qy 239 HKVTEFIDMAAAYAWADVVCRSGLATVSEIAAAGLPALFVPFQHKDQOYVNALPLEK 298
Db 241 YIPVPYVDRMDLAYAADMMCLCAGMTVAELSAVGLPAAVYPLPIGNGEQRINACPVVK 300
Qy 299 AGAAKIIQPOLSVDAVANTLAGWSRETL-----TWAERARAASIPDATERVANE-- 349
Db 301 AGGGLLVDDAELTPE-----WVQGNVLPVLADPHRLYENSRAS--EFGRDADLL 350
Qy 350 VSRVAPALEHHH 361
Db 351 VGMVYERIAARH 362

RESULT 17
Q8GBAO PRELIMINARY; PRT; 379 AA.
AC Q8GBAO;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(Pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.-) (Fragment).
OS Helicobacillus mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacillus.
OX NCBI_TaxID=28064;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22337798; PubMed=12446909;
RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,
RA Blankenship R.E.;
RT "Whole-genome analysis of photosynthetic prokaryotes.";
RL Science 298:1616-1620(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Liolios K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukarenko V.,
RA Gerdes S., Kyrpides N., Overbeek R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AX142807; AAN87411.1; -.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_tran_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR006009; MurG.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRfam; TIGR01133; murG; 1.
KW Transferase; Glycosyltransferase.
FT NON_TER 379 379
SQ SEQUENCE 379 AA; 41095 MW; 4C2EB832148045EA CRC64;

Query Match 23.8%; Score 448.5; DB 2; Length 379;
Best Local Similarity 32.8%; Pred. No. 7.7e-24;
Matches 126; Conservative 71; Mismatches 138; Indels 49; Gaps 14;

Qy 7 KRLVMAGGTGGHVPGLVAHHLMAQ--GWQVRLGTADRMADLVPHKGIIDFIRIS 64
Db 9 RKFVLGGGTGGHVIYALAIARGLQERFPFGCSIEYIGGRGLTIVPREGPLTKVHCR 68
Qy 65 GL-RGKGIKALIAAPLRIENAR--QARAIKAYKDPVLMGVGVYSGPGLAAWSLGI 120
Db 69 GLERGLSLKNIILAA----IGTGRGLLEALIFRLKLPDAVIGTGGEVAFVPMVAATLLGI 124
Qy 121 PVVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAE-----VVGNFVRVTDVLAIP 171

```
Db 125 PALIHEQNAVGVNRLAPRVGSMVLM---TFPEACKRLKANDVIVTGLFVRPSILS-- 178
Qy 172 LPQORLAGR-----EGPVRLVVGSGQARILNQTMPQVAAKLG--DSVLIWHOSGKG 222
Db 179 --ASREGRFPQIPPEAQV--LLVVGSGRGAKLNEAMAPLARNLAGQRYQVHLVHTGES 235
Qy 223 SQSQVEQAYAEAGOPQ-----KVTEFIDDMAAAYAWADVVCRCGALTIVSEIAAAGLPA 277
Db 236 NYDETRLLYEDAGIALEKCGNIKLLPYLDRMDVALAASDLVCVRAGAAFISEITARGLAS 295
Qy 278 LPVDFQK-KRQOQVWNAIPLEKGAAXIIIEQPQLS-----VDVANTLAGWSRETLITMAE 332
Db 296 ILIPYAAENHOEANARSLESTGATKVIDLRELTSQLQEMVPHLLS--HROQVEAMAQ 353
Qy 333 RARAASIPDATERVANEVSRVARA 356
Db 354 AARSAGRPEALAHILSIERWKS 377

RESULT 18
Q8CY50 PRELIMINARY; PRT; 393 AA.
AC Q8CY50;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DE UDP-N-acetylglucosamine--N-acetylmutamyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR BL1323.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
EX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Fessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Fridmore R.D., Arigoni F.,
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RT Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AE014760; AAN25123.1;
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRFAMs; TIGR01133; murG; 1.
KW Complete proteome.
SQ SEQUENCE 393 AA; 41511 MW; DAF16113P5923AAD CRC64;

Query Match 22.9%; Score 431.5; DB 16; Length 393;
Best Local Similarity 32.9%; Pred. No. 1.3e-22;
Matches 129; Conservative 60; Mismatches 156; Indels 47; Gaps 8;

Qy 5 QGKRLMVA--GGTGGHVPFGLVAHHL--MAQGVQVRLGTADRMADLVPKHGIEIDFI 61
Db 3 QGTPHVLVAGGTAGHVNPLAVAGAIRDIEPTAVTIGTAGVLEKDLVEAGVELDTI 62
Qy 62 RISLGRKGKIALIAPRIFNWQRAIMKAYKPDVVLGNGGVVSGFGLAWSGLIP 121
Db 63 EKVPFPRRLNLYRFPFAKKEKATKVRISILETRHADVVAGGVYASAPVYATAKNGIP 122
Qy 122 VVLEHQNGAGTNTKWLARIATKV-----MQAEFGAPFNAEVGNPVRTDVLALPLPQ 175
```

```
Db 123 TAIHEQNARAGMANKLGAHWADFTGVYEGTGLKPRAGADVVERVGLPLRPAIASL---TK 179
Qy 176 RLAGEGEPVR-----VLVVGSGQARILNQTMPQVAAKLGDSVLIWHOSG 220
Db 180 RIGDRAAVRRESAAQLGVDPNRLPLVLTGSLGAQSLNRAIASSAADLLAHAQIIHLTG 239
Qy 221 KGSQSQVEQAY-----EAGOPQHKVTEFIDDMAAAYAWADVVCRCGALTIVSE 266
Db 240 RGIKISEVRELVTASAGADVLTGIGPESAGQDHYHTAEYLERIDILAFACADLVICRAGAGS 299
Qy 267 VSEIAAAGLPAIFVPFQHKRQOQVWNAIPLEKGAAXIIIEQ-----PQLSDVAVANTLAGW 322
Db 300 VSELAALGLPAIYVPLPIGNGEQRFNAEPVNVNAGGLLVADKDLTPQWVHEHVPDLA-- 357
Qy 323 SRETLTWAERARAASIPDATERVANEVSRVA 354
Db 358 DHERLAEFGRKAWEXGIRNAAEIMARHVLQLA 389

RESULT 19
Q83HK1 PRELIMINARY; PRT; 356 AA.
AC Q83HK1;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (EC 2.4.1.1).
GN MURG OR TW542.
OS Tropheryma whipplei (strain TW08/27) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Cellulomonadaceae; Tropheryma.
OX NCBI_TaxID=218496;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22495039; PubMed=12606174;
RA Bentley S.D., Maitwald M., Murphy L.D., Pallen M.J., Yeats C.A.,
RA Dover L.G., Norbertczak H.T., Besta G.S., Quail M.A., Harris D.E.,
RA von Herbay A., Goble A., Rutter S., Squares R., Squares S.,
RA Barrell B.G., Parkhill J., Relman D.A.;
RT "Sequencing and analysis of the genome of the Whipple's disease
RT bacterium tropheryma whipplei."
RT Lancet 361:637-644(2003).
DR EMBL; BX251411; CAD67208.1;
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR000524; HTH_Gntr.
DR InterPro; IPR006009; MurG.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRFAMs; TIGR01133; murG; 1.
DR PROSITE; PS00043; HTH GNTR FAMILY; 1.
DR Transferase; Glycosyltransferase; Complete proteome.
KW Transferase.
SQ SEQUENCE 356 AA; 38420 MW; B5S5608CB471ED9D CRC64;

Query Match 22.9%; Score 430; DB 16; Length 356;
Best Local Similarity 33.1%; Pred. No. 1.5e-22;
Matches 121; Conservative 63; Mismatches 150; Indels 32; Gaps 10;

Qy 8 RLNVMAAGTGGHVPFGLVAHHLMAQGVQVRLGTADRMADLVPKHGIEIDFIRISGI- 66
Db 3 RILLAGGTAGHVNPLALADLVKVGSHATFATGSEGIESRLVNSG--IDFFIPKLP 60
Qy 67 -RGKIGALIAAPLIRIFNWRQARIMKAYKPDVVLGNGGVYSGFGLAWSGLIPVLH 125
```

Db 61 FPRTRSHILCFPFKFFSSVKLVRSILIEHKIQVVVGGYVAAAPAYAAASLNIPYVH 120
Qy 126 EQNGIAGLTNKLARIATKVMQBPQAFPAEVAEYVGNPVRTDVLAL-----PLPQORLAGRE 181
Db 121 ESNARPGLANLLAAHFAKCVGISVIGALPCGKLVGTPIRRDLTAAASFDPVLAKEKGLD 180
Qy 182 GPVR--VLVVGGSGGARILNQTW---POVAAKLGDSVIIW---HOSGKGSQOSVEQAYA 232
Db 181 -PVRKLLLVFGSGGSQSAKINNMHRAALPRVLKLCDEKNYLQVLIHITGYDS-----I 232
Qy 233 EAGOPQHKVTEFIDDMAAAYADVVVCRSGALTVEIAAAGLPALFPFQHKDRQOYWN 292
Db 233 DVNMPHYSSVRVYMDSMGYALSAADLVVSRAGSSTVAELCTFGIPAIYIYPFPGNGEQRN 292
Qy 293 ALPLEKAGAAKIIQOPQLSVDVAVANTLAGWSRETLTMAERARAASIPDATERVANEVSR 352
Db 293 VSHME--SAARIIGENDLSQIRLEDELL-----ELMTDDERREAMSIAAKRPAICNAQN 345
Qy 353 VARALE 358
Db 346 TASLIE 351

RESULT 20

Q820Y4 PRELIMINARY; PRT; 356 AA.
AC Q820Y4;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.-)
GN MURG OR TWT228.
OS Tropheryma whipplei (strain Twist) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Cellulomonadaceae; Tropheryma.
OX NCBI_TaxID=203267;
RN (1)
RP SEQUENCE FROM N.A.
RA Racult D., Audic S., Robert C., Ogata H., Suhre K., Drancourt M.,
EA Claverie J.-M.,
RT Tropheryma whipplei illustrates the diversity of gene loss patterns
in small genome bacterial pathogens.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AE016851; AAC44325.1; -
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_trans_28_C.
DR InterPro; IPR000524; HTH_GntR.
DR InterPro; IPR006009; MurG.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_transf_28_C; 1.
DR TIGRFAMs; TIGR01133; murG; 1.
DR PROSITE; PS00043; HTH_GNTR_FAMILY; 1.
KW Transferase; Glycosyltransferase; Complete proteome.
SQ SEQUENCE 356 AA; 38420 MW; B555608CB471ED9D CRC64;

Query Match 22.9%; Score 430; DB 16; Length 356;
Best Local Similarity 33.1%; Pred. No. 1.5e-22;
Matches 121; Conservative 63; Mismatches 150; Indels 32; Gaps 10;
Qy 8 RLWVAGGTGGVFFGLAVAHHLMAQGNQVRLGTADRMEADLVPKGIEIDFIRISGL- 66
Db 3 RILLAGGGTAGHVNPLALADVLKVSCHATFALGTSEGIESRLVNSG--IDFTTILPLP 60

Search completed: June 7, 2004, 07:15:18
Job time : 51 secs